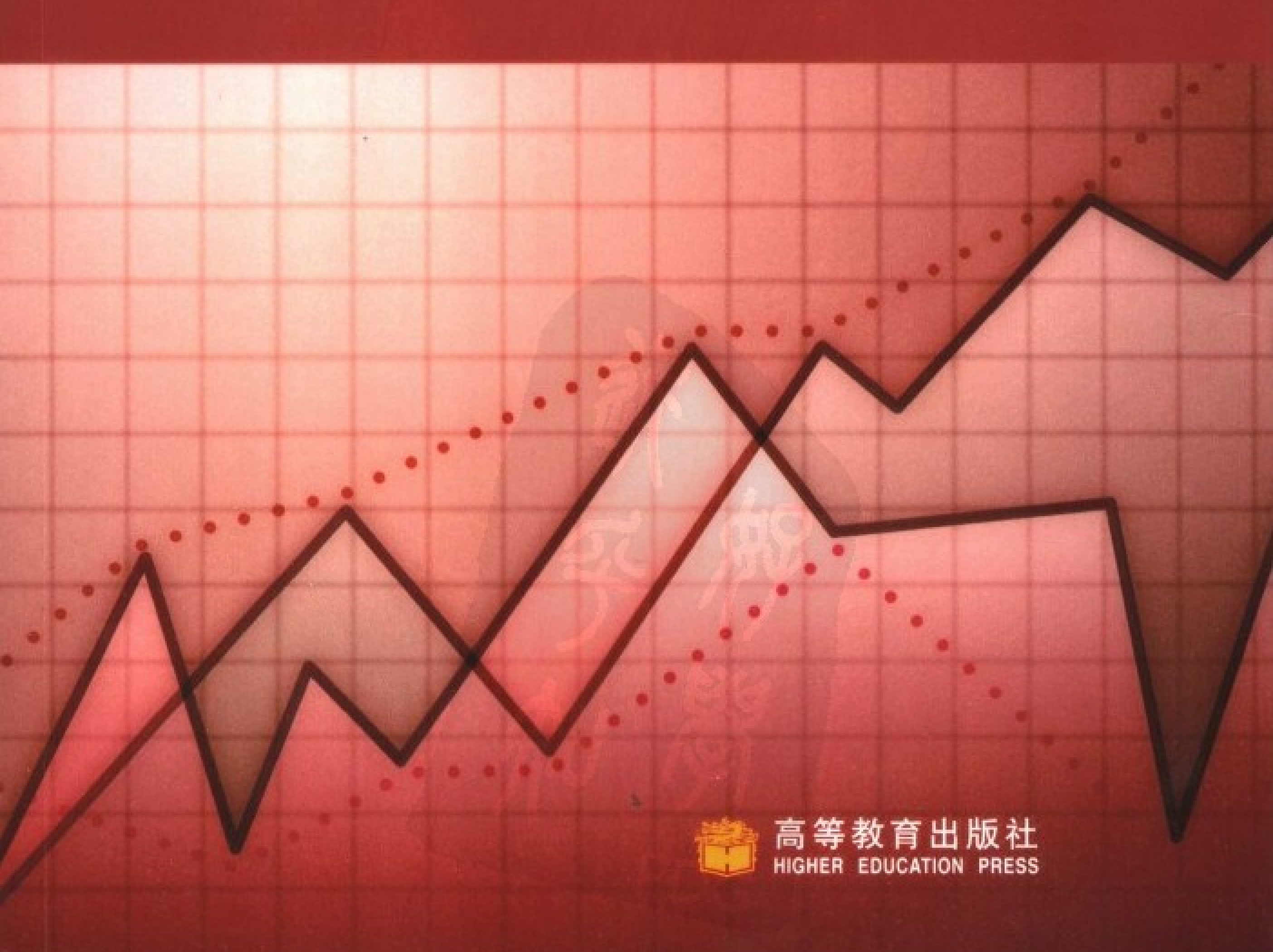
精算学中的随机过程

□ 张连增



高等教育出版社
HIGHER EDUCATION PRESS

本书不同于传统的理工或者经管类的随机过程教科书。在系统介绍了现代精算学中的随机过程理论的基础上，本书将随机过程理论及其在金融保险中的应用有机地结合起来，深入研究出现于金融保险中的随机过程专题，系统揭示随机过程的理论与方法如何巧妙地应用于金融保险中。

本书可作为综合大学经济类、金融类、保险类高年级本科生和研究生的教材或参考书，也可以供保险业精算人员和其他对金融工程、保险精算有兴趣的读者参考。


ISBN 7-04-020457-6

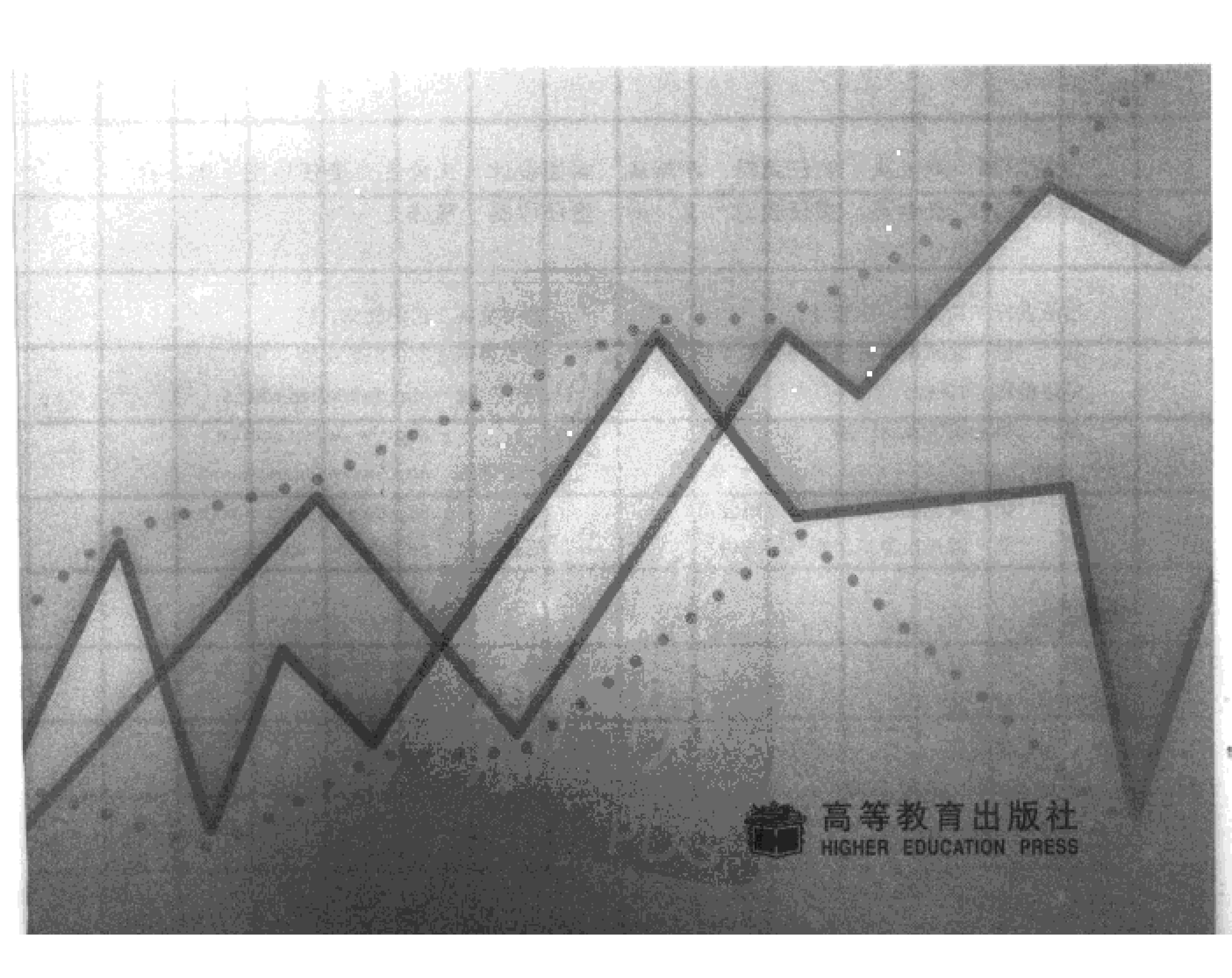
9 787040 204575 >

定价 30.10 元

学科类别：数学

精算学中的随机过程

□ 张连增



高等教育出版社
HIGHER EDUCATION PRESS

图书在版编目（CIP）数据

精算学中的随机过程/张连增. —北京:高等教育出版社,2006.12
ISBN 7-04-020457-6

Ⅰ.精… Ⅱ.张… Ⅲ.精算学-随机过程-高等学校-教材 Ⅳ.F224.0

中国版本图书馆 CIP 数据核字(2006)第 122898 号

策划编辑 赵天夫 责任编辑 邓雁城 封面设计 王凌波 责任绘图 郝 林
版式设计 史新薇 责任校对 王 雨 责任印制 陈伟光

出版发行	高等教育出版社	购书热线	010-58581118
社 址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经 销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印 刷	涿州市星河印刷有限公司	畅想教育	http://www.widedu.com
开 本	787×960 1/16	版 次	2006年12月第1版
印 张	14.5	印 次	2006年12月第1次印刷
字 数	260 000	定 价	30.10元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

物料号 20457-00



前　言

像所有学科一样，精算学也是不断发展的学科。十年前，国际上精算师资格考试的内容较少涉及随机过程，但近年来，随机过程已经逐渐成为国际精算师资格考试的重要内容，随机过程的思想已经体现于精算学的教学研究中，并逐渐体现于保险监管中。在精算学的学术研究方面，浏览一下近年来国际精算学术期刊发表的研究论文，可见随机过程理论与方法已经成为不可缺少的工具。

一方面，应用随机过程理论不但可以处理传统的寿险与非寿险精算问题，而且可以处理非传统的保险产品定价等问题；另一方面，金融与投资理论，尤其是数理金融理论已经成为精算学的重要组成部分，在数理金融理论中，随机过程是最基本的工具。本书以精算学中的随机过程为中心，在以上两个方面加以展开，结合精算学系统地介绍随机过程。因此，本书区别于通常的供数学类等理科师生参考的随机过程的书。在本书中，数学处理上的严谨性是必要的，而且有必要在高起点上组织内容，但还要把随机过程如何在精算学中得到应用这一思路贯穿其中。

本书选材丰富，包含了精算学中常见的随机过程内容。国际上有种观点认为，Markov 过程、鞅、平稳独立增量过程构成了随机过程的绝大部分内容。我同意这种观点，在内容取舍上，以上内容在本书中都有所涉及。关于随机过程理论的一些经典内容，我参考了过去十多年来出版的部分优秀外文专著教材，而一些较新的内容，散见于近年来在国际学术期刊发表的研究论文。

在本书的内容安排上，前三章内容比较直观，读者学过概率论的一些基础知识后，就可以很容易阅读。有了对随机过程的一些感性认识后，从第四章开始，就需要在高起点上对随机过程理论与方法有深入的理解，为此需要概率的公理化定义。按

照我的体会，条件期望是一个核心概念，对初学者也是一个难点，不易理解。但只要完全理解了这个核心概念后，学习后面的涉及鞅、Markov 过程、平稳独立增量过程的内容，就变得非常容易。相对而言，最后一章内容（更新过程）比较独立，但这部分内容在精算学中，尤其是非寿险精算的破产理论会涉及，所以此书包括了更新过程的内容。另外，各章内容相对比较独立，便于读者阅读某个专题。

需要指出，本书中某些章节的内容没有完全展开，或者对某些结论没有给出证明，例如鞅的收敛定理、可选停时定理、随机微分方程、线性偏微分方程的数值算法等。这出于两方面的考虑。一方面，我希望在选材上安排好结构与要点，另一方面，从介绍随机过程如何在精算学中得到应用的观点，把某些内容完全展开意义不是很大。读者在精算学的学习与研究中，如需要进一步对某些专题有深入的了解，可参考相关的优秀专著教材。

十多年前，我在南开大学数学系学习随机过程，其后在南开大学风险管理与保险学系任精算学教师。十多年来，我对随机过程的理解逐渐深入，同时见证了随机过程理论如何逐渐深入地在精算学中得以应用。我一直希望以自己的所学所悟，完成一本专著，力争成为国内精算学优秀参考书，能够对广大读者有所启发帮助，尽快地了解前沿内容。

数年前，我阅读精算学研究论文，已经注意到 Norberg 教授的研究，很具特色。2004 年下半年，我访问了伦敦经济学院统计学系，更深入地了解了 Norberg 教授主持的精算学项目。此前，我于 2002 年访问了墨尔本大学经济学系精算研究中心，了解了该中心 Dickson 教授主持的精算学项目。对国内外精算学术研究的差别，我深有体会，也非常希望在进行学术研究的同时，为国内同行撰写一本专著，把国际上最新的研究成果系统地介绍给国内学术界同行，一方面供同行参考，另一方面为广大的莘莘学子铺路。

我于 2004 年冬季和 2005 年冬季两次为南开大学风险管理与保险学系精算研究生开设随机过程课程。本书是在讲学内容基础上扩充整理、反复修改而来，凝聚了我的艰苦努力与许多心血。我尽了最大的努力，系统地整理总结多年来的所学所悟。撰写本书的难度，比我原来设想的要大得多。但艰辛的付出总有回报，我深深地体会到，本书的撰写过程也是一个不断加深理解的过程。

借此机会，我要对许多人表达我深深的谢意，没有他们方方面面的支持与帮助，这本书是不可能出现的。首先，我要感谢家人的大力支持。感谢南开大学风险管理与保险学系 2004 级和 2005 级的精算研究生，通过他们的反馈，我更容易知道本书的难点所在，这对未来的教学很有帮助。我特别感谢 2005 级研究生李开顺、高云、王子辉、周琪、魏建等人，其中李开顺、高云、王子辉分别帮助完成了第一章、第二章、第三章的初稿，周琪帮助完成了第四章的部分初稿，魏建帮助绘制了书中的插图。另外，我要感谢 Norberg 教授和 Dickson 教授，他们为我提供了对外学术交流

的机会。

最后,本书能够如期出版,我要感谢高等教育出版社的大力支持。每一本书的出版都凝聚了多位编辑的辛勤劳动，本书也不例外。我特别要感谢两位编辑赵天夫和邓雁城，他们细致地审阅了全书，指出了一些内容和形式的不当之处，给出了一些内行的改进建议，从而使本书增色不少。

张连增

2006 年 6 月 18 日



目　录



第一章　离散时间Markov 链

§1　转移概率与 Chapman-Kolmogorov 方程

1. 定义与例子

这一章要讨论的随机过程 $\{X_n, n=0,1,2,\cdots\}$, 状态空间包含有限个或可列个值。除非另有说明, 在这一章中, 状态空间是非负整数集 $\{0,1,2,\cdots\}$。$\{X_n=i\}$ 表示过程在时刻 n 处于状态 i 这一事件。假设过程当前处于状态 i, 下一步它以概率 P_{ij} 转移到另一状态 j, 而且 P_{ij} 与此前的信息无关, 即对于所有状态 $i_0, i_1, \cdots, i_{n-1}, i, j$, 以及 $n \geqslant 0$,

$$\mathbb{P}\{X_{n+1}=j|X_n=i, X_{n-1}=i_{n-1}, \cdots, X_1=i_1, X_0=i_0\}=P_{ij} \tag{1}$$

此时称 X 为(时间齐次)Markov链。(1)式表明, 对于 Markov 链, 在给定过去状态 $X_0, X_1, \cdots, X_{n-1}$ 和现在状态 X_n 的条件下, 将来状态 X_{n+1} 的条件分布只取决于现在状态, 而与过去状态无关。

由概率的非负性, 考虑到各种转移方式, 可得

$$P_{ij} \geqslant 0, \quad i,j \geqslant 0; \quad \sum_{j=0}^{\infty} P_{ij}=1, \quad i=0,1,\cdots$$

以 $\mathbf{P}$ 表示一步转移概率 P_{ij} 构成的矩阵,

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

下面给出一些 Markov 链的例子。

例 1 (天气预报) 假设明天会下雨的概率只取决于今天的天气状态, 而与过去的天气状态无关。如果今天下雨, 那么明天下雨的概率为 α; 如果今天不下雨, 那么明天下雨的概率为 β。

如果用状态 0 表示下雨, 用状态 1 表示不下雨, 那么可以构造两种状态的 Markov 链, 其转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

例 2 (信息系统) 考虑一个输送数字 0 和 1 的信息系统。每个数字的输送必须经过若干个阶段。在每个阶段, 一个数字在离开时与进入时保持不变的概率为 p。设 X_n 表示数字在第 n 个阶段内的变化, 那么 $\{X_n, n = 0, 1, 2, \cdots\}$ 是两种状态的 Markov 链, 其转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

例 3 在一天内, 人的心情或者高兴(A), 或者一般(B), 或者忧郁(C)。如果今天是高兴的, 那么明天的心情是 A、B 或 C 的概率分别为 0.5, 0.4, 0.1。如果今天感觉一般, 那么明天的心情是 A、B 或 C 的概率分别为 0.3, 0.4, 0.3。如果今天是忧郁的, 那么明天的心情是 A、B 或 C 的概率分别为 0.2, 0.3, 0.5。

设 X_n 表示在第 n 天的心情, 那么 $\{X_n, n \geqslant 0\}$ 是三种状态的 Markov 链, 其转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

例 4 (把一个过程转化为 Markov 链) 假设今天是否下雨取决于过去两天的天气条件。具体地讲, 如果过去两天都下雨, 那么明天下雨的概率为 0.7; 如果今天下雨而昨天没下雨, 那么明天下雨的概率为 0.5; 如果昨天下雨而今天没下雨, 那么明天下雨的概率为 0.4; 如果过去两天都没下雨, 那么明天下雨的概率为 0.2。

设时刻 n 的状态只取决于在时刻 n 是否下雨, 那么上述模型不是 Markov 链。但是, 如果任意时刻的状态由当天和前一天的天气条件决定, 那么上述模型可以转化为 Markov 链。此时, 共有以下四种状态:

状态 0　如果今天和昨天都下雨

状态 1　如果今天下雨而昨天没下雨

状态 2　如果昨天下雨而今天没下雨

状态 3　如果今天和昨天都没下雨

此时可构造 Markov 链, 其转移矩阵为

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}$$

例 5 (随机游动模型) 如果 $P_{i,i+1} = p, P_{i,i-1} = 1-p,\ i = 0, \pm 1, \pm 2, \cdots$, 其中 $0 < p < 1$, 那么称状态空间是全体整数 $\{0, \pm 1, \pm 2, \cdots\}$ 的 Markov 链为随机游动。

上述模型的直观解释如下: 一个人沿一条直线行走, 在每一时刻, 他或者以概率 p 向右移动一步, 或者以概率 $1-p$ 向左移动一步。

例 6 (赌博模型) 假设一个赌徒在每一盘赌局中, 或者以概率 p 赢得 1 元, 或者以概率 $1-p$ 输掉 1 元。进一步, 假设赌徒在输光或者在赢到 N 元时退出赌局, 那么该赌徒的财富变化构成 Markov 链, 其转移概率为

$$P_{i,i+1} = p = 1 - P_{i,i-1}, i = 1, 2, \cdots, N-1$$

$$P_{00} = P_{NN} = 1$$

这里称状态 0 和 N 为吸收状态, 即过程进入就停留在那里的状态。上述例子是一个有限状态的有吸收壁(状态 0 和 N)的随机游动。

2. Chapman-Kolmogorov 方程

前面定义了一步转移概率 P_{ij}。现在定义过程从状态 i 经过 n 步转移到状态 j 的 n 步转移概率 P_{ij}^n,

$$P_{ij}^n = \mathbb{P}\{X_{m+n} = j | X_m = j\}, \quad i, j \geqslant 0$$

P_{ij}^1 就是前面的 P_{ij}。为计算 n 步转移概率, 可应用如下 Chapman-Kolmogorov 方程: 对所有的 $n, m \geqslant 0$, 以及状态 i, j,

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m \tag{2}$$

直观地讲, (2)式右边的 $P_{ik}^n P_{kj}^m$ 表示过程从初始状态 i 出发, 先在第 n 步转移到状态 k, 再经过 m 步转移到状态 j 的概率。考虑所有可能的中间状态 k, 就得到(2)式。(2)式的证明如下:

$$\begin{aligned}
P_{ij}^{n+m} &= \mathbb{P}\{X_{n+m} = j | X_0 = i\} \\
&= \sum_{k=0}^{\infty} \mathbb{P}\{X_{n+m} = j, X_n = k | X_0 = i\} \\
&= \sum_{k=0}^{\infty} \mathbb{P}\{X_n = k | X_0 = i\} \mathbb{P}\{X_{n+m} = j | X_n = k, X_0 = i\} \\
&= \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m
\end{aligned}$$

设 $\mathbf{P}^{(n)}$ 表示n步转移概率 P_{ij}^n 构成的矩阵, 那么由(2)式可得

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \mathbf{P}^{(m)}$$

特别地,

$$\begin{aligned}
\mathbf{P}^{(2)} &= \mathbf{P}^{(1+1)} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 \\
\mathbf{P}^{(n)} &= \mathbf{P}^{(n-1+1)} = \mathbf{P}^{(n-1)}\mathbf{P} = \mathbf{P}^n
\end{aligned}$$

即 n 步转移概率矩阵等于一步转移概率矩阵$\mathbf{P}$的 n 次幂。

例 7 在例 1 中, 天气变化构成两种状态的 Markov 链。如果 $\alpha = 0.7$, $\beta = 0.4$, 计算在今天下雨的条件下 4 天后下雨的概率。

已知一步转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

下面计算 $\mathbf{P}^{(4)}$,

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix}$$

$$\mathbf{P}^{(4)} = (\mathbf{P}^2)^2 = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix}$$

最后, 所求的概率 $P_{00}^4 = 0.5749$。

例 8 在例 4 中, 已知星期一和星期二下雨, 求在星期四下雨的概率。

已知二步转移概率矩阵为

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix} \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.49 & 0.12 & 0.21 & 0.18 \\ 0.35 & 0.20 & 0.15 & 0.30 \\ 0.20 & 0.12 & 0.20 & 0.48 \\ 0.10 & 0.16 & 0.10 & 0.64 \end{bmatrix}$$

在星期四下雨等价于过程处于状态 0 或者状态 1, 因此所求的概率为

$$P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61$$

至此, 所考虑的概率都是条件概率。例如,P_{ij}^n 表示在时刻 0 时过程处于状态 i, 而后在时刻 n 过程处于状态 j 的概率。如果要确定在时刻 n 的状态的无条件分布, 那么需要在时刻 0 的状态的概率分布。为此引入符号

$$\alpha_i = \mathbb{P}\{X_0 = i\},\ i \geqslant 0$$

那么无条件概率就可以计算如下,

$$\mathbb{P}\{X_n = j\} = \sum_{i=0}^{\infty} \mathbb{P}\{X_n = j | X_0 = i\} \mathbb{P}\{X_0 = i\} = \sum_{i=0}^{\infty} P_{ij}^n \alpha_i$$

在例 7 中, 如果 $\alpha_0 = 0.4,\ \alpha_1 = 0.6$, 那么在 4 天后会下雨的无条件概率为

$$\begin{aligned} \mathbb{P}\{X_4 = 0\} &= 0.4P_{00}^4 + 0.6P_{10}^4 \\ &= 0.4 \times 0.5749 + 0.6 \times 0.5668 = 0.5700 \end{aligned}$$

§2 状态分类

根据不同的分类准则, 可对 Markov 链中的状态进行分类。

1. 相通状态

如果存在 $n \geqslant 0$, 使得 $P_{ij}^n > 0$, 那么就称状态 i 可到达状态 j。相反地, 如果不存在这样的 n, 就称状态 i 不能到达状态 j。如果两种状态 i 和 j 是彼此可到达的, 则称状态 i 和状态 j 相通, 记为 $i \leftrightarrow j$。

每个状态与自身是相通的, 这是因为根据定义,

$$P_{ii}^0 = \mathbb{P}\{X_0 = i | X_0 = i\} = 1$$

相通关系有以下三种性质:

(i) 对于所有的 $i \geqslant 0$, $i \leftrightarrow i$。

(ii) 如果 $i \leftrightarrow j$, 那么 $j \leftrightarrow i$。

(iii) 如果 $i \leftrightarrow j$, 而且 $j \leftrightarrow k$, 那么 $i \leftrightarrow k$。

性质 (i) 和 (ii) 根据相通定义即可得到。现证明性质(iii) 。由假设 $i \to j$, $j \to k$ 可知, 存在正整数 n 和 m, 使得 $P_{ij}^n > 0$, $P_{jk}^m > 0$。由 Chapman-Kolmogorov 方程可得

$$P_{ik}^{n+m} = \sum_{r=0}^{\infty} P_{ir}^n P_{rk}^m \geqslant P_{ij}^n P_{jk}^m > 0$$

所以, 状态 i 可达状态 k, 同样可证状态 k 可达状态 i。因此 $i \leftrightarrow k$。

两个相通的状态属于同一个类。由性质(i) 、(ii) 、(iii) 可以很容易得到下述结论: 根据相通概念, 可以把状态空间分成不相交的子集合, 每个子集合内的任意两个状态都是相通的。如果 Markov 链的状态空间只包含一个子集合, 即任意两个状态都是相通的, 那么就称该 Markov 链是不可约的。

例 9 考虑由 0,1,2 三个状态构成的 Markov 链, 其转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.25 & 0.25 \\ 0 & 1/3 & 2/3 \end{bmatrix}$$

容易证明此 Markov 链是不可约的。例如, 从状态 0 到达状态 2 是可能的, 因为

$$0 \to 1 \to 2$$

即从状态 0 到达状态 2 的一条路径是以 0.5的概率从状态 0 到达状态 1, 再以 0.25 的概率从状态 1 到达状态 2。

再考虑由 0,1,2,3 四个状态构成的另一个 Markov 链, 其转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

此 Markov 链包含三个类{0,1}、{2}、{3}。状态 2 可到达状态 0(或 1), 而状态 0(或 1)不能到达状态 2。状态 3 是一个吸收状态, $P_{33}=1$, 所以它不能到达其他状态。

2. 常返状态与非常返状态

对于任一状态 i, 设 f_i 为过程从状态 i 出发, 在未来某一时刻再次返回状态 i 的概率。如果 $f_i=1$, 那么称状态i是常返的, 如果 $f_i<1$, 那么称状态 i 是非常返的。

假设过程从状态 i 出发, 而且状态 i 是常返的, 那么过程以概率 1 将再次回到状态 i。根据 Markov 链的定义, 当过程再次回到状态 i 时, 它会再次从 i 出发, 所以状态 i 会再次到达。重复这个推理, 就会得出结论: 如果状态 i 是常返的, 那么从状态 i 出发, 过程将会无限多次地回到状态 i。

相对应地, 假设状态 i 是非常返的, 那么过程从状态 i 出发, 将会以正概率 $1-f_i$ 不再回到状态 i。因此, 从状态 i 出发, 过程正好 n 次回到状态 i 的概率等于 $f_i^n(1-f_i), n\geqslant 0$。换言之, 如果状态 i 是非常返的, 那么从状态 i 出发, 过程回到状态i的次数服从一个均值为 $f_i/(1-f_i)$ 的几何分布。

总之, 如果状态 i 是常返的, 当且仅当从状态 i 出发, 过程回到状态 i 的次数是无限多的。

引入符号

$$I_n=\begin{cases}1, & X_n=i\\ 0, & X_n\neq i\end{cases}$$

那么 $\sum\limits_{n=0}^{\infty} I_n$ 表示过程处于状态 i 的次数。

$$\begin{aligned}\mathbb{E}\left[\sum_{n=0}^{\infty} I_n|X_0=i\right]&=\sum_{n=0}^{\infty}\mathbb{E}[I_n|X_0=i]\\&=\sum_{n=0}^{\infty}\mathbb{P}\{X_n=i|X_0=i\}=\sum_{n=0}^{\infty}P_{ii}^n\end{aligned}$$

总结前面的结论, 即得以下性质。

性质 1 (i) 状态 i 是常返的当且仅当 $\sum\limits_{n=0}^{\infty}P_{ii}^n=\infty$。

(ii) 状态i是非常返的当且仅当 $\sum\limits_{n=0}^{\infty}P_{ii}^n<\infty$。

有限状态的 Markov 链必存在常返状态。为了说明这点, 假设状态空间为 $\{0,1,\cdots,M\}$, 而且所有状态都是非常返的。设过程的初始状态为 0, 那么经过有限次

之后, 过程不再回到状态 0。设过程从一个新的状态出发, 那么经过有限次之后, 过程不再回到此新状态。重复这一推理, 那么经过有限次之后, 所有的状态都将不会被访问。但是, 实际上过程总处于某种状态, 这表明至少存在一种常返状态。

下面性质表明常返状态可构成相通类。

性质 2 如果状态 i 是常返的, 并且状态 i 与状态 j 相通, 那么状态 j 是常返的。

证明 因为状态 i 与状态 j 相通, 所以存在正整数 k 与 m, 使得 $P_{ij}^k > 0, P_{ji}^m > 0$。另外, 对于任意正整数 n, 以下不等式成立

$$P_{jj}^{m+n+k} \geqslant P_{ji}^m P_{ii}^n P_{ij}^k$$

上述不等式的左边表示在 $m+n+k$ 步内状态 j 返回状态 j 的概率, 而不等式右边表示状态 j 先在 m 步内到达状态 i, 其后状态 i 在另外的 n 步内返回状态 i, 然后状态 i 在另外的 k 步内到达状态 j 这样一条特殊路径在 $m+n+k$ 步内返回状态 j 的概率。

对上述不等式关于 n 求和, 由性质 1 可得

$$\sum_{n=1}^{\infty} P_{jj}^{m+n+k} \geqslant P_{ji}^m P_{ij}^k \sum_{n=1}^{\infty} P_{ii}^n = \infty$$

再次应用性质 1(i) 可得状态 j 也是常返的。

性质 2 表明非常返状态和常返状态不属于同一相通类。即如果状态 i 是非常返的, 且与状态 j 相通, 那么状态 j 也是非常返的。这是因为如果状态 j 是常返的, 那么根据性质 2, 状态 i 也是常返的。

如前所述, 在有限状态的 Markov 链中, 总存在常返状态。结合性质 2, 可得出关于有限状态的不可约的 Markov 链的一个结论: 所有的状态都是常返的。

例 10 设由状态 0,1,2,3 构成的 Markov 链的转移概率矩阵为

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

容易验证所有的状态都是相通的, 因此所有的状态都是常返的。

例 11 设 Markov 链的状态空间为 $\{0,1,2,3,4\}$, 转移概率矩阵为

$$\mathbf{P}=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

状态空间可分为三个类: $\{0,1\},\{2,3\}$ 和 $\{4\}$, 其中前 2 个类是常返的, 第 3 个类是非常返的。

3. 随机游动

例 12 (随机游动) Markov 链的状态空间为全体整数 $\{0,\pm1,\pm2,\cdots\}$, 转移概率为

$$P_{i,i+1}=p=1-P_{i,i-1},\quad 0<p<1$$

形象地说, 随机游动可描述为一个醉汉在一条路上左右行走。另外, 随机游动也可以描述一个赌徒在每盘赌局中赢一元或输一元, 从而导致财富变化的过程。

因为所有状态都是相通的, 根据性质 2 可得, 所有的状态或者都是常返的, 或者都是非常返的。因此只需考虑状态 0, 为此考虑 $\sum\limits_{n=1}^{\infty}P_{00}^{n}$。

设初始状态为 0, 那么经过奇数次转移后, 状态不会回到 0, 因此

$$P_{00}^{2n-1}=0,\quad n=1,2,\cdots$$

另一方面, 在赌博模型下, 经过了 $2n$ 次赌博后, 当且仅当赢了 n 局, 同时也输了 n 局, 才能输赢相等。每盘赌局赢的概率为 p, 输的概率为 $1-p$, 所以

$$P_{00}^{2n}=\begin{pmatrix}2n\\n\end{pmatrix}p^{n}(1-p)^{n}=\frac{(2n)!}{n!n!}[p(1-p)]^{n},n=1,2,3,\cdots$$

应用 Stirling 公式, 当 n 很大时,

$$n!\approx n^{n+(1/2)}\mathrm{e}^{-n}\sqrt{2\pi} \tag{3}$$

把(3)式代入 P_{00}^{2n} 的表达式, 可得

$$P_{00}^{2n}\approx\frac{[4p(1-p)]^{n}}{\sqrt{\pi n}}$$

因此, 当且仅当

$$\sum_{n=1}^{\infty}\frac{[4p(1-p)]^n}{\sqrt{\pi n}}$$

收敛时, $\sum_{n=1}^{\infty}P_{00}^n$ 收敛。注意到 $4p(1-p)\leqslant 1$, 而且仅当 $p=1/2$ 时, 等号成立。因此当 $p=1/2$ 时 $\sum_{n=1}^{\infty}P_{00}^n=\infty$, 状态 0 是常返的。而当 $p\neq 1/2$ 时,$\sum_{n=1}^{\infty}P_{00}^n<\infty$, 状态 0 是非常返的。

当 $p=1/2$ 时, 相应的随机游动称为对称随机游动。在多维空间里也有对称随机游动。例如, 在二维对称随机游动中, 设过程在每次移动中将以 1/4 的概率向左、右、上、下移动一步。此时每个状态是整数对 (i,j), 转移概率为

$$P_{(i,j),(i+1,j)}=P_{(i,j),(i-1,j)}=P_{(i,j),(i,j+1)}=P_{(i,j),(i,j-1)}=\frac{1}{4}$$

下面应用与一维情形中相似的方法, 证明二维对称随机游动的所有状态都是常返的。因为二维对称随机游动是不可约的, 只需证明状态 0=(0,0) 是常返的。

考虑 P_{00}^{2n}。如果在前 $2n$ 步移动中, 左移 i 步, 右移 i 步, 上移 $n-i$ 步, 下移 $n-i$ 步, 其中 $i\leqslant n$, 那么过程将会返回到初始状态。因为每步移动是四种类型中的一种, 而且出现每种类型的概率均为 1/4, 所以所求的概率是一个多项分布概率, 即

$$\begin{aligned}P_{00}^{2n}&=\sum_{i=0}^{n}\frac{(2n)!}{i!i!(n-i)!(n-i)!}\left(\frac{1}{4}\right)^{2n}\\&=\sum_{i=0}^{n}\frac{(2n)!}{n!n!}\frac{n!}{(n-i)!i!}\frac{n!}{(n-i)!i!}\left(\frac{1}{4}\right)^{2n}\\&=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}\sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i}\\&=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}\binom{2n}{n}\end{aligned}\tag{4}$$

上面的最后一个等式用到了以下组合性质

$$\binom{2n}{n}=\sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i}$$

上述等式两边均表示从含有 n 个白球和 n 个黑球的盒子中抽取 n 个球, 所有的抽取方式数目。应用 Stirling 公式, 当 n 很大时,

$$\binom{2n}{n}=\frac{(2n)!}{n!n!}\approx\frac{(2n)^{2n+(1/2)}\mathrm{e}^{-2n}\sqrt{2\pi}}{n^{2n+1}\mathrm{e}^{-2n}(2\pi)}=\frac{4^n}{\sqrt{\pi n}}$$

代入(4)式, 最后得到

$$P_{00}^{2n} \approx \frac{1}{\pi n}$$

由此 $\sum_n P_{00}^{2n} = \infty$, 所以状态 0, 从而所有的状态都是常返的。

有趣的是, 虽然一维和二维对称随机游动都是常返的, 但所有其他的高维对称随机游动都是非常返的, 具体证明这里不再展开。(例如, 三维对称随机游动每次等可能地转移到六种状态中的一种 —— 左、右、上、下、前、后。)

现在回到一维随机游动, 设初始状态为 0, 当 $p \neq 1/2$ 时, 求过程在未来某一时刻再次回到状态 0 的概率 f_0。首先注意到, 经过一步转移后, 应用全概率公式,

$$f_0 = \mathbb{P}\{\text{在某一时刻返回}|X_1 = 1\}p + \mathbb{P}\{\text{在某一时刻返回}|X_1 = -1\}(1-p)$$

假设 $p > 1/2$, 那么 $\mathbb{P}\{\text{在某一时刻返回}|X_1 = -1\} = 1$, 所以

$$f_0 = \mathbb{P}\{\text{在某一时刻返回}|X_1 = 1\}p + 1 - p \tag{5}$$

记 $\alpha = \mathbb{P}\{\text{在某一时刻返回}|X_1 = 1\}$, 考虑下一步转移后的状态, 得到

$$\begin{aligned}\alpha &= \mathbb{P}\{\text{在某一时刻返回}|X_1 = 1, X_2 = 0\}(1-p) \\ &\quad + \mathbb{P}\{\text{在某一时刻返回}|X_1 = 1, X_2 = 2\}p \\ &= 1 - p + \mathbb{P}\{\text{在某一时刻返回}|X_0 = 2\}p\end{aligned}$$

如果过程处于状态 2, 那么为了回到状态 0, 它必须首先回到状态 1, 而且这一事件在某一时刻发生的概率也是 α。另外, 如果过程到达了状态 1, 那么它在另一时刻到达状态 0 的概率也是 α。所以, 从状态 2 出发, 在未来某一时刻到达状态 0 的概率是 α^2。总之,

$$\alpha = 1 - p + p\alpha^2$$

上述方程的两个根是 $\alpha = 1$ 和 $\alpha = (1-p)/p$。根据非常返性, $\alpha < 1$, 因此, $\alpha = (1-p)/p$。代入(5)式, 可得

$$f_0 = 1 - p + 1 - p = 2(1-p)$$

类似地, 当 $p < 1/2$ 时, 可以证明 $f_0 = 2p$。总之, 一般地

$$f_0 = 2\min\{p, 1-p\}$$

4. 一个应用例子

例 13 (关于 Aloha 协议的最大不稳定性) 在每一时期 $n=1,2,\cdots$ 到达一个通信设备的信息数是相互独立同分布的随机变量。记 $a_i=\mathbb{P}\{i\text{个信息到达}\}$, 假设 $a_0+a_1<1$。每一到达的信息将会在期末输送。如果只有一条信息被输送, 那么信息就会成功地离开系统。但是, 如果在同一时刻有两条或多条信息需要输送, 那么必定会发生冲突, 使这些信息保留在系统中。发生冲突的每条信息独立于所有其他信息, 在下一时期末以概率 p 被输送, 这就是所谓的 Aloha 协议。下面证明这个系统是渐近不稳定的, 即成功输送的次数以概率 1 是有限的。

以 X_n 表示在第 n 个时期开始时设备中的信息数, $\{X_n,n\geqslant 0\}$ 是 Markov 链。对于 $k\geqslant 0$, 定义指示变量 I_k, 其中

$$I_k=\begin{cases}1, & \text{Markov 链首次离开状态}k\text{时, 直接进入状态}k-1\\ 0, & \text{其他}\end{cases}$$

如果系统从未处于状态 $k,k\geqslant 0$, 那么设 $I_k=0$。(例如, 如果相邻的状态是 $0,1,3,3,4,\cdots$, 那么 $I_3=0$, 因为过程第一次离开状态 3 时, 它进入了状态 4; 而如果相邻的状态是 $0,3,3,2,\cdots$, 那么 $I_3=1$, 因为这次它进入了状态 2。)

$$\begin{aligned}\mathbb{E}\left[\sum_{k=0}^{\infty}I_k\right]&=\sum_{k=0}^{\infty}\mathbb{E}[I_k]=\sum_{k=0}^{\infty}\mathbb{P}\{I_k=1\}\\&\leqslant\sum_{k=0}^{\infty}\mathbb{P}\{I_k=1|k\text{曾被访问过}\}\end{aligned}\tag{6}$$

$\mathbb{P}\{I_k=1|k\text{曾被访问过}\}$ 表示当过程离开状态 k, 下一步到达状态 $k-1$ 的概率。它是一个条件概率, 即在过程不能返回到状态 k 的条件下, 从状态 k 转移到状态 $k-1$ 的概率,

$$\mathbb{P}\{I_k=1|k\text{曾被访问过}\}=\frac{P_{k,k-1}}{1-P_{kk}}$$

因为

$$P_{k,k-1}=a_0\begin{pmatrix}k\\1\end{pmatrix}p(1-p)^{k-1}=a_0kp(1-p)^{k-1}$$
$$P_{k,k}=a_0[1-kp(1-p)^{k-1}]+a_1(1-p)^k$$

以上两式可这样理解: (i) 如果某天开始有 k 条信息, 当天没有新的信息到达, 而且 k 条信息中的一条信息被输送, 那么在第二天开始将会有 $k-1$ 条信息; (ii) 如果某天开始有 k 条信息, 或者当天没有新信息到达, 也没有任何信息被输送, 或者当天正好有一条新信息到达(它自动被输送), 那么在第二天开始会有 k 条信息。

把上面两式代入(6)式, 可得

$$\mathbb{E}\left[\sum_{k=0}^{\infty} I_k\right] \leqslant \sum_{k=0}^{\infty} \frac{a_0 kp(1-p)^{k-1}}{1-a_0[1-kp(1-p)^{k-1}]-a_1(1-p)^k} < \infty$$

为说明上式的收敛性, 注意到当 k 充分大时, 求和项的分母趋近于 $1-a_0$, 所以求和式是收敛的等价于 $\sum_{k=0}^{\infty} k(1-p)^{k-1} < \infty$, 而此不等式是成立的。

$\mathbb{E}\left[\sum_{k=0}^{\infty} I_k\right] < \infty$ 表明 $\sum_{k=0}^{\infty} I_k < \infty$ 以概率 1 成立。否则, 如果存在一个正的概率, 使得 $\mathbb{P}\left\{\sum_{k=0}^{\infty} I_k = \infty\right\} = \delta > 0$, 那么 $\mathbb{E}\left[\sum_{k=0}^{\infty} I_k\right] = \infty$。因此, 信息成功输送而离开的状态数以概率 1 是有限的。等价地说, 存在一个有限正整数 N, 使得当系统中有 N 条或更多信息时, 将不会再有成功的输送。从这点出发, 并注意到状态数总会最终到达很大的事实, 可以得出, 有限多次成功输送的概率为 1。

5. Stirling 公式

最后给出关于 Stirling 公式的一个粗略的概率证明。设 $X_1, X_2, \cdots$ 是相互独立同分布的 Poisson 随机变量, 均值为 1。设 $S_n = \sum_{i=1}^{n} X_i$, 那么 S_n 的均值和方差均为 n。于是

$$\begin{aligned}
\mathbb{P}\{S_n = n\} &= \mathbb{P}\{n-1 < S_n \leqslant n\} \\
&= \mathbb{P}\left\{\frac{-1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leqslant 0\right\} \\
&\approx \int_{\frac{-1}{\sqrt{n}}}^{0} (2\pi)^{\frac{-1}{2}} \mathrm{e}^{\frac{-x^2}{2}} \mathrm{d}x \\
&\approx (2\pi)^{\frac{-1}{2}} \left(\frac{1}{\sqrt{n}}\right) = (2\pi n)^{\frac{-1}{2}}
\end{aligned}$$

在上述推导中, 第三式应用了中心极限定理, 第四步用到了中值定理做近似。

另一方面, S_n 服从均值为 n 的 Poisson 分布, 所以

$$\mathbb{P}\{S_n = n\} = \frac{n^n}{n!}\mathrm{e}^{-n}$$

因此, 当 n 充分大时,

$$\frac{n^n}{n!}\mathrm{e}^{-n} \approx (2\pi n)^{\frac{-1}{2}}$$

变形就得 Stirling 公式

$$n! \approx n^{n+\frac{1}{2}}\mathrm{e}^{-n}\sqrt{2\pi}$$

§3 极限概率

1. 极限概率

在例 7 中, 对于两种状态的 Markov 链, $\mathbf{P}^{(4)}$ 的数值结果为

$$\mathbf{P}^{(4)} = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix}$$

由此得到 $\mathbf{P}^{(8)} = \mathbf{P}^{(4)}\mathbf{P}^{(4)}$(保留三位有效数字)

$$\mathbf{P}^{(8)} = \begin{bmatrix} 0.572 & 0.428 \\ 0.570 & 0.430 \end{bmatrix}$$

注意到矩阵 $\mathbf{P}^{(8)}$ 与矩阵 $\mathbf{P}^{(4)}$ 几乎相同, 而且 $\mathbf{P}^{(8)}$ 的各行几乎都相同。实际上, 对于所有的 i, 当 $n \to \infty$ 时, P_{ij}^n 均收敛于某个值, 而且极限与初始状态 i 无关。直观上讲, 过程经过许多步转移之后, 处于状态 j 的概率趋近于极限概率, 而且极限值与初始状态无关。

为了更准确地分析以上观察, 需考虑关于 Markov 链的状态的另外两个性质。对满足 $P_{ii}^n > 0$ 的正整数集 $\{n\}$, 设 d 是最大公约数, 那么称状态 i 的周期为 d。例如, 从状态 i 出发, 过程回到状态 i 所需的步数为 $2, 4, 6, 8, \cdots$, 那么状态 i 的周期就是 2。周期为 1 的状态称为非周期的。可以证明, 如果状态 i 的周期为 d, 而且 i 与 j 相通, 那么状态 j 的周期也为 d。

设状态 i 是常返的, 进一步假设从状态 i 出发返回状态 i 的转移步数的期望是有限的, 就称状态 i 为正常返的。如果从状态 i 出发返回状态 i 的转移步数的期望是无限的, 那么就称状态 i 为零常返的。可以证明, 如果状态 i 与 j 相通, 而且状态 i 是正常返的, 那么状态 j 也是正常返的。虽然存在零常返状态, 但可以证明对有限状态的 Markov 链, 所有的常返状态都是正常返的。非周期的正常返状态称为遍历状态。

以下不加证明地给出一个重要定理。对有限状态的 Markov 链, 可应用矩阵论的结论来证明以下定理。

定理 对于不可约的遍历的 Markov 链, 对所有的 i, j, $\lim\limits_{n\to\infty} P_{ij}^n$ 存在, 而且极限与 i 无关。记

$$\pi_j = \lim_{n\to\infty} P_{ij}^n, \quad j \geqslant 0$$

那么 π_j 是方程组

$$\begin{aligned} \pi_j &= \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j \geqslant 0 \\ \sum_{i=0}^{\infty} \pi_j &= 1 \end{aligned} \tag{7}$$

的唯一非负解。

下面给出一些注记。

(i) 如果已知 $\pi_j = \lim\limits_{n\to\infty} P_{ij}^n$ 存在, 而且极限与初始状态 i 无关, 不难得出 π_j 必满足(7)式, 只需注意到 $\lim\limits_{n\to\infty} \mathbf{P}^{(n)} = \mathbf{P}\left(\lim\limits_{n\to\infty} \mathbf{P}^{(n)}\right)$ 即可。或者可以从另一角度出发, 考虑时刻 n 时的各个状态, 可以得到 $\mathbb{P}\{X_{n+1}=j\}$ 的如下表达式:

$$\begin{aligned}\mathbb{P}\{X_{n+1}=j\} &= \sum_{i=0}^{\infty} \mathbb{P}\{X_{n+1}=j|X_n=i\}\mathbb{P}\{X_n=i\} \\ &= \sum_{i=0}^{\infty} P_{ij}\mathbb{P}\{X_n=i\}\end{aligned}$$

设 $n\to\infty$, 假设可以交换求和与极限顺序, 那么

$$\pi_j = \sum_{i=0}^{\infty} P_{ij}\pi_i$$

(ii) 可以证明, 过程在时刻 n 处于状态 j 的极限概率 π_j 等于在长时期内, 过程处于状态 j 的次数比例。

(iii) 对不可约的、正常返的、有周期的 Markov 链, 以下方程

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad j \geqslant 0$$

$$\sum_{i=0}^{\infty} \pi_j = 1$$

仍有唯一非负解。但是此时 π_j 只能解释为在长时期内, Markov 链处于状态 j 的次数比例。

2. 一些例子

例 14 在例 1 中, 如果今天下雨, 那么明天下雨的概率为 α; 如果今天不下雨, 那么明天下雨的概率为 β。如果以状态 0 表示下雨, 状态 1 表示不下雨, 那么由(7)式, 极限概率 π_0, π_1 为以下方程的解

$$\pi_0 = \alpha\pi_0 + \beta\pi_1$$

$$\pi_1 = (1-\alpha)\pi_0 + (1-\beta)\pi_1$$

$$\pi_0 + \pi_1 = 1$$

求解得到

$$\pi_0 = \frac{\beta}{1+\beta-\alpha}, \quad \pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

特别地, 设 $\alpha = 0.7, \beta = 0.4$, 那么下雨的极限概率为 $\pi_0 = \dfrac{4}{7} = 0.571$。

例 15　在例 3 中, 一个人的心情变化构成三种状态的 Markov 链, 转移概率矩阵为

$$\mathbf{P}=\begin{bmatrix}0.5 & 0.4 & 0.1\\ 0.3 & 0.4 & 0.3\\ 0.2 & 0.3 & 0.5\end{bmatrix}$$

求在长时期内, 过程处于各种状态的次数比例。

求解由(7)式构成的方程组。在本例中, 这些式子为

$$\pi_0=0.5\pi_0+0.3\pi_1+0.2\pi_2$$

$$\pi_1=0.4\pi_0+0.4\pi_1+0.3\pi_2$$

$$\pi_2=0.1\pi_0+0.3\pi_1+0.5\pi_2$$

$$\pi_0+\pi_1+\pi_2=1$$

解得

$$\pi_0=\frac{21}{62},\quad \pi_1=\frac{23}{62},\quad \pi_2=\frac{18}{62}$$

例 16 (阶层流动模型)　社会学家感兴趣的一个问题是确定社会上从事上层或下层职业的人口比例。作为一个可行的数学模型, 假设一个家族中连续几代之间的社会阶层转变是 Markov 链状态的转移, 即假设子女的职业仅取决于其父母的职业。假设转移概率矩阵为

$$\mathbf{P}=\begin{bmatrix}0.45 & 0.48 & 0.07\\ 0.05 & 0.70 & 0.25\\ 0.01 & 0.50 & 0.49\end{bmatrix}\tag{8}$$

例如, 假设一个中层的孩子将会从事上层、中层或下层职业的概率分别为 0.05, 0.70, 0.25。

极限概率 π_i 满足

$$\pi_0=0.45\pi_0+0.05\pi_1+0.01\pi_2$$

$$\pi_1=0.48\pi_0+0.70\pi_1+0.50\pi_2$$

$$\pi_2=0.07\pi_0+0.25\pi_1+0.49\pi_2$$

$$\pi_0+\pi_1+\pi_2=1$$

求解得到

$$\pi_0=0.07,\quad \pi_1=0.62,\quad \pi_2=0.31$$

换言之, 一个社会的阶层流动可用转移概率矩阵为(8)的 Markov 链表示。在长时期内, 有 7% 的人从事上层工作, 有 62% 的人从事中层工作, 有 31% 的人从事下层工作。

例 17 (遗传学中的 Hardy-Weinberg 法则与 Markov 链) 考虑一个有大量个体的群体, 每个个体都拥有一对特殊的基因, 每个基因或为 A 型, 或为 a 型。假设基因对是 AA, aa 和 Aa 的个体比例分别为 $p_0, q_0, r_0 (p_0 + q_0 + r_0 = 1)$。当两个个体配对时, 每个个体随机提供各自基因中的一个, 产生后代。假设配对也是随机发生的, 每一个体都等可能地与另一个体配对, 我们感兴趣的是求出下一代基因对分别为 AA, aa 和 Aa 的个体比例, 设这些比例为 p, q, r。通过关注下一代个体, 就可以很容易求出这些比例。

先随机选择一个父代, 再随机选择一个基因, 等价于从基因总体中选择一个基因。以父代的基因对为条件, 可以得到随机选择的一个基因是 A 型的概率为

$$\begin{aligned}\mathbb{P}\{\mathrm{A}\} &= \mathbb{P}\{\mathrm{A}|\mathrm{AA}\}p_0 + \mathbb{P}\{\mathrm{A}|\mathrm{aa}\}q_0 + \mathbb{P}\{\mathrm{A}|\mathrm{Aa}\}r_0 \\ &= p_0 + r_0/2\end{aligned}$$

类似地, 基因是 a 型的概率为

$$\mathbb{P}\{\mathrm{a}\} = q_0 + r_0/2$$

所以, 在随机配对下, 随机选择的下一代成员是 AA 型的概率为 p, 其中

$$p = \mathbb{P}\{\mathrm{A}\}\mathbb{P}\{\mathrm{A}\} = (p_0 + r_0/2)^2$$

类似地, 随机选择的下一代成员是 aa 型的概率为

$$q = \mathbb{P}\{\mathrm{a}\}\mathbb{P}\{\mathrm{a}\} = (q_0 + r_0/2)^2$$

下一代成员是 Aa 型的概率

$$r = 2\mathbb{P}\{\mathrm{A}\}\mathbb{P}\{\mathrm{a}\} = 2(p_0 + r_0/2)(q_0 + r_0/2)$$

既然下一代的每个成员都独立地以概率 p, q, r 分别取 AA, aa 和 Aa 型基因, 因此下一代成员中是 AA, aa 和 Aa 型的比例分别为 p, q, r。

如果考虑下一代的总基因库, 那么 A 型基因的比例 $p + r/2$ 与父代相比不会发生改变, 这是因为代与代之间的总基因库不会发生改变, 它也可由下面的简单代数运算得到,

$$\begin{aligned}p + r/2 &= (p_0 + r_0/2)^2 + (p_0 + r_0/2)(q_0 + r_0/2) \\ &= (p_0 + r_0/2)[p_0 + r_0/2 + q_0 + r_0/2] \\ &= p_0 + r_0/2 \\ &= \mathbb{P}\{\mathrm{A}\}\end{aligned} \tag{9}$$

在(9)式的推导中, 用到 $p_0+q_0+r_0=1$。由此即知, 下一代基因库中 A 型和 a 型的比例与父代相应的比例相同。因此, 在随机配对的条件下, 所有初始代以后的相继各代中, 群体中 AA 型, aa 型和 Aa 型基因对的比例会保持在 p,q,r 上。这就是著名的 Hardy-Weinberg 法则。

假设总体基因对的比例稳定地保持在 p,q,r 上。现在考察单个个体及其后代的基因历史。为了简便, 假设每个个体只有一个后代。对于一个给定的个体, 以 X_n 表示他的第 n 代后代的基因状态。在随机选择配对的条件下, X 构成 Markov 链, 它的转移概率矩阵为

$$\begin{array}{c} \\ \text{AA} \\ \text{aa} \\ \text{Aa} \end{array} \begin{array}{c} \begin{array}{ccc} \text{AA} & \text{aa} & \text{Aa} \end{array} \\ \begin{bmatrix} p+\dfrac{r}{2} & 0 & q+\dfrac{r}{2} \\ 0 & q+\dfrac{r}{2} & p+\dfrac{r}{2} \\ \dfrac{p}{2}+\dfrac{r}{4} & \dfrac{q}{2}+\dfrac{r}{4} & \dfrac{p}{2}+\dfrac{q}{2}+\dfrac{r}{2} \end{bmatrix} \end{array}$$

此 Markov 链的极限概率(个体后代处于三种基因状态的比例)应该正好是 p,q,r, 这是非常直观的。为了验证这一结论, 需证明 p,q,r 满足(7)式。因为(7)式中的一个方程是多余的, 只需证明以下三式即可。

$$p=p(p+r/2)+r(p/2+r/4)=(p+r/2)^2$$

$$q=q(q+r/2)+r(q/2+r/4)=(q+r/2)^2$$

$$p+q+r=1$$

由前述分析, 上述三式是成立的。

例 18　假设一个生产流程的状态转变可用 Markov 链来描述, 转移概率为 $P_{ij}, i, j=1,\cdots,n$。假设有些状态是可接受的, 而另外一些状态是不可接受的。以 A 表示可接受状态, A^c 表示不可接受状态。生产流程处于可接受状态时称它为"上升" 的, 反之称它为"下降"的, 当生产流程持续很长时间后,确定以下量:

(i) 生产流程从上升到下降的转换率。

(ii) 当生产流程下降时, 它保持在下降状态的平均时间。

(iii) 当生产流程上升时, 它保持在上升状态的平均时间。

下面分别考虑各个问题。

(i) 设 $\{\pi_k, k=1,\cdots,n\}$ 表示极限概率分布。于是对于 $i\in A$ 和 $j\in A^c$, 过程从状态 i 进入状态 j 的转换率为 $\pi_i P_{ij}$。所以生产流程从可接受状态进入状态 j 的转换率为 $\sum\limits_{i\in A}\pi_i P_{ij}$, 最后从可接受状态进入不可接受状态的转换率为

$$\sum_{j\in A^c}\sum_{i\in A}\pi_i P_{ij} \tag{10}$$

(ii) 和(iii) 以 $\bar{U}$ 表示当生产流程上升时保持上升的平均时间。相应地，以 $\bar{D}$ 表示保持下降的平均时间。因为平均来说，在 $\bar{U}+\bar{D}$ 段时间内生产流程会有一次损坏，所以单位时间内生产流程的损坏率为

$$\frac{1}{\bar{U}+\bar{D}}$$

结合(10)式，可得关于 $\bar{U}$ 和 $\bar{D}$ 的一个关系

$$\frac{1}{\bar{U}+\bar{D}}=\sum_{j\in A^c}\sum_{i\in A}\pi_i P_{ij} \tag{11}$$

为了得到关于 $\bar{U}$ 和 $\bar{D}$ 的另一关系，考虑生产流程处于上升状态的时间比例，它等于 $\sum\limits_{i\in A}\pi_i$。另一方面，由于在 $\bar{U}+\bar{D}$ 段时间内，生产流程平均有 $\bar{U}$ 段时间处于上升状态，所以

$$\frac{\bar{U}}{\bar{U}+\bar{D}}=\sum_{i\in A}\pi_i \tag{12}$$

由(11)式和(12)式，可得

$$\bar{U}=\frac{\sum\limits_{i\in A}\pi_i}{\sum\limits_{j\in A^c}\sum\limits_{i\in A}\pi_i P_{ij}}$$

$$\bar{D}=\frac{1-\sum\limits_{i\in A}\pi_i}{\sum\limits_{j\in A^c}\sum\limits_{i\in A}\pi_i P_{ij}}=\frac{\sum\limits_{i\in A^c}\pi_i}{\sum\limits_{j\in A^c}\sum\limits_{i\in A}\pi_i P_{ij}}$$

例如，设 Markov 链的状态空间为 $\{1,2,3,4\}$，转移概率矩阵为

$$\mathbf{P}=\begin{bmatrix}1/4 & 1/4 & 1/2 & 0\\ 0 & 1/4 & 1/2 & 1/4\\ 1/4 & 1/4 & 1/4 & 1/4\\ 1/4 & 1/4 & 0 & 1/2\end{bmatrix}$$

设可接受的(上升)状态为 1, 2，不可接受的(下降)状态为 3, 4。极限概率满足以下线性方程组，

$$\pi_1=\frac{1}{4}\pi_1+\frac{1}{4}\pi_3+\frac{1}{4}\pi_4$$

$$\pi_2=\frac{1}{4}\pi_1+\frac{1}{4}\pi_2+\frac{1}{4}\pi_3+\frac{1}{4}\pi_4$$

$$\pi_3 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{4}\pi_3$$

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

解得

$$\pi_1 = \frac{3}{16}, \quad \pi_2 = \frac{1}{4}, \quad \pi_3 = \frac{14}{48}, \quad \pi_4 = \frac{13}{48}$$

所以单位时间内生产流程的损坏率为

$$\pi_1(P_{13} + P_{14}) + \pi_2(P_{23} + P_{24}) = \frac{9}{32} \approx 0.28$$

而平均持续时间为

$$\bar{U} = \frac{14}{9} \approx 1.56, \quad \bar{D} = 2$$

因此, 单位时间内生产流程平均有 28% 的时间发生损坏。发生损坏后, 平均持续 2 个时间单位, 其后, 系统上升, 平均持续 1.56 个时间单位。

3. 平稳分布

下面给出关于 π_j 的两个注记。

(i) $\{\pi_j, j \geqslant 0\}$ 经常被称为平稳概率分布。这是因为, 如果初始状态的概率分布为 $\{\pi_j, j \geqslant 0\}$, 那么在任意时刻 n 过程处于状态 j 的概率就等于 π_j。具体地说, 如果

$$\mathbb{P}\{X_0 = j\} = \pi_j, \quad j \geqslant 0$$

那么对所有的 n,

$$\mathbb{P}\{X_n = j\} = \pi_j$$

上述结论由归纳法容易证明。假设对于 $n-1$ 上式成立, 那么

$$\begin{aligned}\mathbb{P}\{X_n = j\} &= \sum_i \mathbb{P}\{X_n = j | X_{n-1} = i\}\mathbb{P}\{X_{n-1} = i\} \\ &= \sum_i P_{ij}\pi_i = \pi_j\end{aligned}$$

(ii) 对状态 j, 以 m_{jj} 表示 Markov 链从状态 j 出发到首次返回到状态 j 的步长的期望。因为平均来说, 在 m_{jj} 个时间单位内, 有 1 个时间单位过程处于状态 j, 所以

$$\pi_j = \frac{1}{m_{jj}}$$

总之, 过程处于状态 j 的时间比例等于从 j 出发首次回到 j 的平均时间的倒数。实际上, 此结论是更新过程理论中一般结论的一个特例。

例 19 考虑一枚硬币的独立抛掷问题, 在每次抛掷中以概率 p 出现正面(简记为 H), 以概率 $q=1-p$ 出现反面(简记为 T)。求出现 HTHT 模式所需的抛掷次数的期望。

为解答这个问题, 假设出现 HTHT 模式后, 硬币抛掷仍将无限制地继续下去。当 $n \geqslant 4$ 时, 定义时刻 n 的状态是过去最近的 4 次结果, 当 $n \leqslant 3$ 时, 定义时刻 n 的状态是过去最近的 n 次结果。那么可以看出相继的状态构成 Markov 链。例如, 如果前 5 次结果是 TTHH, 那么 Markov 链的相继状态为 $X_1=\mathrm{T}, X_2=\mathrm{TT}, X_3=\mathrm{TTT}, X_4=\mathrm{TTTH}, X_5=\mathrm{TTHH}$。根据前面的注记(ii) , 状态 HTHT 的极限概率 π_{HTHT} 等于从状态 HTHT 首次回到状态 HTHT 的平均时间的倒数。对于任意的 $n \geqslant 4$, 为使时刻 n 的状态是 HTHT, 第 n 次抛掷结果须为 T, 第 $n-1$ 次须为 H, 第 $n-2$ 次须为 T, 第 $n-3$ 次须为 H。因为各次抛掷是相互独立的, 所以

$$\mathbb{P}\{X_n=\mathrm{HTHT}\}=p^2q^2, \quad n \geqslant 4$$

$$\pi_{\mathrm{HTHT}}=\lim_{n\to\infty}\mathbb{P}\{X_n=\mathrm{HTHT}\}=p^2q^2$$

因此, $1/(p^2q^2)$ 是从状态 HTHT 首次回到状态 HTHT 的平均时间。这意味着从状态 HT 出发, 首次到达状态 HTHT 的平均时间也是 $1/(p^2q^2)$。为了到达 HTHT, 首先要到达 HT, 所以

$$\mathbb{E}[\text{到达HTHT的时间}]=\mathbb{E}[\text{到达HT的时间}]+\frac{1}{p^2q^2}$$

为了求出上式右边的第一项, 应用同样的方法, 区别在于定义时刻 n 的状态为过去最近的两次结果。由此得到从状态 HT 首次回到状态 HT 的平均时间为 $1/(pq)$, 这也是状态 HT 首次出现所需要的平均时间。从而

$$\mathbb{E}[\text{到达HTHT的时间}]=\frac{1}{pq}+\frac{1}{p^2q^2}$$

应用同样的方法, 可以得到任意给定模式出现所需的平均时间。例如,

$$\begin{aligned}\mathbb{E}[\text{HTHHTHTHH出现所需的时间}]&=\mathbb{E}[\text{HTHH出现所需的时间}]+\frac{1}{p^6q^3}\\&=\mathbb{E}[\text{H出现所需的时间}]+\frac{1}{p^3q}+\frac{1}{p^6q^3}\\&=\frac{1}{p}+\frac{1}{p^3q}+\frac{1}{p^6q^3}\end{aligned}$$

另外, 基本试验不必局限于只有两种可能结果(如 H 和 T)。例如, 如果相继结果是独立同分布的, 以 p_j 表示取值等于 j 的概率, 那么

$$\begin{aligned}\mathbb{E}[\text{出现012301所需的时间}]&=\mathbb{E}[\text{出现01所需的时间}]+\frac{1}{p_0^2p_1^2p_2p_3}\\&=\frac{1}{p_0p_1}+\frac{1}{p_0^2p_1^2p_2p_3}\end{aligned}$$

在本节最后，给出以下性质。

性质 3　设 $\{X_n, n \geqslant 1\}$ 是 Markov 链，平稳分布为 $\{\pi_j, j \geqslant 0\}$，r 是定义于状态空间的一个有界函数，那么下式以概率 1 成立

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N} r(X_n)}{N} = \sum_{j=0}^{\infty} r(j)\pi_j$$

证明　如果以 $a_j(N)$ 表示 Markov 链在前面 N 个时期 $1, \cdots, N$ 处于状态 j 的时期数，那么

$$\sum_{n=1}^{N} r(X_n) = \sum_{j=0}^{\infty} a_j(N) r(j)$$

上式两边同除以 N，再令 $N \to \infty$，注意到 $a_j(N)/N \to \pi_j$，就得性质 3。

如果 $r(j)$ 表示 Markov 链处于状态 j 时对应的回报，那么性质 3 表明单位时间内平均回报为 $\sum_j r(j)\pi_j$。

§4　赌徒破产问题及其在药物试验中的应用

1. 赌徒破产问题

设赌徒在每盘赌局中以概率 p 赢得 1 单位，以概率 $q = 1 - p$ 输掉 1 单位。假设相继的赌局是相互独立的，设赌徒的初始财富为 i 单位，求赌徒的财富在到达 0 之前先到达 N 的概率。

以 X_n 表示赌徒在时刻 n 的财富，那么过程 $\{X_n, n = 0, 1, 2, \cdots\}$ 是 Markov 链，其转移概率为

$$P_{00} = P_{NN} = 1$$

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \qquad i = 1, 2, \cdots, N-1$$

Markov 链的状态空间可分为三个类：$\{0\}, \{1, 2, \cdots, N-1\}, \{N\}$；第一类和第三类是常返的，第二类是非常返的。因为每个非常返状态只能被有限次访问，所以经过有限时间之后，赌徒或者实现他的目标 N，或者破产。

对 $i = 0, 1, \cdots, N$，以 P_i 表示从 i 开始，赌徒的财富最终到达 N 的概率。考虑到首次赌博结果，可得

$$P_i = pP_{i+1} + qP_{i-1}, \qquad i = 1, 2, \cdots, N-1$$

因为 $p+q=1$, 上式可变形为

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}), \qquad i = 1, 2, \cdots, N-1$$

注意到 $P_0 = 0$, 由上式得到

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1$$

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

$$\vdots$$

$$P_i - P_{i-1} = \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1$$

$$\vdots$$

$$P_N - P_{N-1} = \frac{q}{p}(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1$$

对前面 $i-1$ 个式子相加, 可得

$$P_i - P_1 = P_1 \left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

等价地

$$P_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)} P_1, & \dfrac{q}{p} \neq 1 \\ iP_1, & \dfrac{q}{p} = 1 \end{cases}$$

再由 $P_N = 1$, 可以得到

$$P_1 = \begin{cases} \dfrac{1-(q/p)}{1-(q/p)^N}, & p \neq \dfrac{1}{2} \\ \dfrac{1}{N}, & p = \dfrac{1}{2} \end{cases}$$

因此

$$P_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^N}, & p \neq \dfrac{1}{2} \\ \dfrac{i}{N}, & p = \dfrac{1}{2} \end{cases} \tag{13}$$

由(13)式可知, 当 $N \to \infty$ 式,

$$P_i \to \begin{cases} 1 - \left(\dfrac{q}{p}\right)^i, & p > \dfrac{1}{2} \\ 0, & p \leqslant \dfrac{1}{2} \end{cases}$$

所以, 当 $p > 1/2$ 时, 赌徒的财富将无限增加的概率为正数; 而当 $p \leqslant 1/2$ 时, 形象地讲, 赌徒就像与一个拥有巨额财富的对手赌博, 必然会破产。

下面给出一个数值例子。

例 20 甲乙两人投硬币, 谁投的硬币最靠近墙壁就获胜, 并赢取对方的硬币。乙是技术高的玩家, 在每次投硬币中赢的概率为 0.6。

(i) 如果在开始时, 甲有 10 枚硬币, 乙有 5 枚硬币, 那么乙将赢光甲的概率是多少?

(ii) 如果在开始时, 甲有 20 枚硬币, 乙有 10 枚硬币, 那么乙将赢光甲的概率是多少?

应用赌徒问题模型, 可以求解如下。

(i) 此时 $i = 5$, $N = 15$, $p = 0.6$, 代入(13)式可得所求概率, 它为

$$\frac{1 - \left(\dfrac{2}{3}\right)^5}{1 - \left(\dfrac{2}{3}\right)^{15}} \approx 0.87$$

(ii) 此时 $i = 10$, $N = 30$, $p = 0.6$, 所求概率为

$$\frac{1 - \left(\dfrac{2}{3}\right)^{10}}{1 - \left(\dfrac{2}{3}\right)^{30}} \approx 0.98$$

2. 赌徒破产问题在药物试验中的应用

假设开发了两种新药用于治疗某种疾病。第 i 号药物的治愈率为 $P_i, i = 1, 2$, 它的含义是每个病人服用第 i 号药物的治愈率为 P_i。这些治愈率是未知的, 我们感兴趣的是能判断 $P_1 > P_2$ 或 $P_2 > P_1$ 的方法。为此, 考虑以下分次试验: 每次试验治疗两个病人, 其中一个病人接受 1 号药物的治疗, 另一个病人接受 2 号药物的治疗。观察每次试验结果, 当一种药物的累积治愈人数超过另一种药物的累积治愈人数, 并且超出的人数达到某个预先给定的数时, 试验就停止。更确切地, 引入

$$X_j = \begin{cases} 1, & \text{如果第}j\text{对病人中接受1号药物治疗的病人病愈} \\ 0, & \text{其他} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{如果第}j\text{对病人中接受2号药物治疗的病人病愈} \\ 0, & \text{其他} \end{cases}$$

对于一个预先给定的正整数 M, 经过 N 次之后试验停止, 其中 N 是使下列条件之一成立的最小的 n。

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = M$$

或者

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = -M$$

在前一种条件下, 可以断定 $P_1 > P_2$, 而在后一种条件下, 可以断定 $P_2 > P_1$。

为了确定上述试验的优良性, 需要知道试验导致错误结论的概率。例如, 对于给定的 P_1 和 P_2, 如果事实上 $P_1 > P_2$, 但试验却导致 $P_2 > P_1$ 的概率。为求这个概率, 注意到每次试验后, 观察服用 1 号药物与 2 号药物的治愈病人, 两者之差的累积值或者以概率 $P_1(1-P_2)$ 上升 1(1 号药物治愈病人而 2 号药物没有治愈病人), 或者以概率 $(1-P_1)P_2$ 下降 1, 或者以概率 $P_1P_2 + (1-P_1)(1-P_2)$ 保持不变。因此, 如果只考虑导致累积值出现差异的试验, 那么累积值上升 1 的概率为

$$p = \mathbb{P}\{\text{上升 1}|\text{上升 1或下降 1}\} = \frac{P_1(1-P_2)}{P_1(1-P_2)+(1-P_1)P_2}$$

而累积值下降 1 的概率为

$$q = 1 - p = \frac{P_2(1-P_1)}{P_1(1-P_2)+(1-P_1)P_2}$$

因此, 试验结果导致 $P_2 > P_1$ 的概率, 等于赢的概率为 p 的赌徒在赢得 M 赌注前输掉 M 赌注的概率。在(13)式中, 令 $i = M, N = 2M$, 可得这个概率

$$\mathbb{P}\{\text{试验导致}P_2 > P_1\} = 1 - \frac{1-(q/p)^M}{1-(q/p)^{2M}} = \frac{1}{1+(q/p)^M}$$

例如, 设真实值为 $P_1 = 0.6, P_2 = 0.4$, 那么当 $M = 5$ 时, 试验导致错误 $P_2 > P_1$ 的概率为 0.017。而当 $M = 10$ 时, 试验导致错误判断的概率减少到 0.0003。

§5 处于非常返状态的平均时间

1. 非常返状态的逗留时间

现考虑有限状态的 Markov 链, 假设状态被编号后, 非常返状态集为 $\mathbf{T} = \{1,$

$2,\cdots,t\}$。记

$$\mathbf{P_T}=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1t} \\ \vdots & \vdots & & \vdots \\ P_{t1} & P_{t2} & \cdots & P_{tt} \end{bmatrix}$$

$\mathbf{P_T}$ 仅表示从非常返状态到非常返状态的转移概率, 所以在 $\mathbf{P_T}$ 中至少存在一行, 使得此行的和小于 1。

对于非常返状态 i 和 j, 以 s_{ij} 表示 Markov 链从状态 i 出发, 到达状态j后并处于状态 j 的时间的期望, 可称为非常返状态的逗留时间。引入 $\delta_{i,j}$, 当 $i=j$ 时, $\delta_{i,j}=1$, 否则 $\delta_{i,j}=0$。考虑到首次转移, 可得

$$s_{ij}=\delta_{i,j}+\sum_{k}P_{ik}s_{kj}=\delta_{i,j}+\sum_{k=1}^{t}P_{ik}s_{kj} \tag{14}$$

因为常返状态构成了封闭子集, 过程不会从常返状态到达非常返状态, 所以(14)式的第二个等式成立。

以 $\mathbf{S}$ 表示 $\{s_{ij},i,j=1,\cdots,t\}$ 构成的矩阵,

$$\mathbf{S}=\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1t} \\ \vdots & \vdots & & \vdots \\ s_{t1} & s_{t2} & \cdots & s_{tt} \end{bmatrix}$$

应用矩阵符号, (14)式可写成

$$\mathbf{S}=\mathbf{I}+\mathbf{P_T}\mathbf{S}$$

其中$\mathbf{I}$ 是t阶单位矩阵。上式等价于

$$(\mathbf{I}-\mathbf{P_T})\mathbf{S}=\mathbf{I}$$

两边同乘以 $(\mathbf{I}-\mathbf{P_T})^{-1}$,

$$\mathbf{S}=(\mathbf{I}-\mathbf{P_T})^{-1}$$

即 $\{s_{ij},i,j\}$ 可以通过求矩阵 $\mathbf{I}-\mathbf{P_T}$ 的逆得到。

例 21 考虑赌徒破产问题, 其中 $p=0.4,N=7$。设初始财富为 3, 计算

(i) 赌徒财富是 5 的次数的期望。

(ii) 赌徒财富是 2 的次数的期望。

首先, 矩阵 $\mathbf{P_T}$ 有如下所示,

$$
\mathbf{P_T} = \begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1 & 0 & 0.4 & 0 & 0 & 0 & 0 \\
2 & 0.6 & 0 & 0.4 & 0 & 0 & 0 \\
3 & 0 & 0.6 & 0 & 0.4 & 0 & 0 \\
4 & 0 & 0 & 0.6 & 0 & 0.4 & 0 \\
5 & 0 & 0 & 0 & 0.6 & 0 & 0.4 \\
6 & 0 & 0 & 0 & 0 & 0.6 & 0
\end{array}
$$

$\mathbf{I}-\mathbf{P_T}$ 的逆为(可应用数学软件直接求出)

$$
\mathbf{S} = (\mathbf{I}-\mathbf{P_T})^{-1} = \begin{bmatrix}
1.6149 & 1.0248 & 0.6314 & 0.3691 & 0.1943 & 0.0777 \\
1.5372 & 2.5619 & 1.5784 & 0.9228 & 0.4857 & 0.1943 \\
1.4206 & 2.3677 & 2.9990 & 1.7533 & 0.9228 & 0.3691 \\
1.2458 & 2.0763 & 2.6299 & 2.9990 & 1.5784 & 0.6314 \\
0.9835 & 1.6391 & 2.0763 & 2.3677 & 2.5619 & 1.0248 \\
0.5901 & 0.9835 & 1.2458 & 1.4206 & 1.5372 & 1.6149
\end{bmatrix}
$$

从以上矩阵, 即得

$$
s_{3,5} = 0.9228, \qquad s_{3,2} = 2.3677
$$

2. 非常返状态的到达概率

对于 $i \in \mathbf{T}, j \in \mathbf{T}$, 以 f_{ij} 表示 Markov 链从状态 i 出发, 在某一时刻进入状态 j 的概率, 称它为非常返状态的到达概率。f_{ij} 通过 $\mathbf{P_T}$ 可以容易求得。为了寻找它们之间的关系, 考虑在某一时刻过程是否到达状态 j, 可得 s_{ij} 的如下表达式。

$$
\begin{aligned}
s_{ij} &= \mathbb{E}[\text{处于}j\text{的时间}|\text{从}i\text{出发, 在某一时刻转移到}j]f_{ij} \\
&\quad + \mathbb{E}[\text{处于}j\text{的时间}|\text{从}i\text{出发, 从未转移到}j](1-f_{ij}) \\
&= (\delta_{ij} + s_{jj})f_{ij} + \delta_{ij}(1-f_{ij}) \\
&= \delta_{ij} + f_{ij}s_{jj}
\end{aligned}
$$

其中 s_{jj} 是从状态 i 出发最终进入状态 j 之后, 处于状态 j 的时间的期望。解上述方程,

$$
f_{ij} = \frac{s_{ij} - \delta_{ij}}{s_{jj}}
$$

例 22 在例 21 中, 赌徒在某时刻的财富为 1 个单位的概率是多少?
由例 21 中的结论 $s_{3,1} = 1.4206, s_{1,1} = 1.6149$, 所以

$$f_{3,1} = \frac{s_{3,1}}{s_{1,1}} = 0.8797$$

作为检验, 注意到 $f_{3,1}$ 是赌徒从 3 开始, 在到达 7 之前到达 1 的概率。它也等于赌徒从 2 开始在到达 6 之前破产的概率。所以

$$f_{3,1} = 1 - \frac{1-(0.6/0.4)^2}{1-(0.6/0.4)^6} = 0.8797$$



第二章　Poisson 过程

§1　Poisson 过程的定义

1. 计数过程

如果以 $N(t)$ 表示在时间区间 $[0,t]$ 内某一特定随机事件发生的总次数, 那么称随机过程 $\{N(t), t \geqslant 0\}$ 为计数过程。由定义可知一个计数过程 $\{N(t), t \geqslant 0\}$ 有以下性质:

(i) $N(t)$ 取值为非负整数。

(ii) 当 $s < t$ 时, $N(s) \leqslant N(t)$, 而且 $N(t) - N(s)$ 表示 $(s,t]$ 时间内事件发生的次数。

下面给出计数过程的一些实例。

例 1　以 $N(t)$ 表示到时刻 t 为止到达某商店的顾客人数, 那么对顾客到达商店这一事件来说, $\{N(t), t \geqslant 0\}$ 是一个计数过程。需要指出, 如果以 $N(t)$ 表示在时刻 t 时商店里的顾客人数, 那么 $\{N(t), t \geqslant 0\}$ 就不是一个计数过程。

例 2　以一个婴儿的出生代表一个事件的发生, 那么以 $N(t)$ 表示到时刻 t 为止的出生总人数时, $\{N(t), t \geqslant 0\}$ 是一个计数过程。

例 3　以 $N(t)$ 表示某个足球运动员到时刻 t 为止已进的球数, 足球运动员踢进一球就表示这个计数过程的事件发生了一次, 那么 $\{N(t), t \geqslant 0\}$ 是一个计数过程。

如果一个计数过程在不相交的时间区间内发生的事件次数是独立的, 则称它具

有独立增量性。此时, 在时刻 s 之前发生的事件次数 $N(s)$ 与在时间 $(s,t]$ 内发生的事件次数 $N(t)-N(s)$ 是独立的。

在例 1 中, 独立增量性的假设有较充分的理由。但在例 2 中, 如果 $N(t)$ 很大, 即在 $[0,t]$ 内出生的婴儿很多, 那么在时刻 t 时存活的人可能很多, 这可能会导致在时间 $(t,t+s]$ 内出生的人数也会很多, 即 $N(t)$ 与 $N(t+s)-N(t)$ 独立的理由不充分。在例 3 中, 在假设足球运动员现在进球的几率不依赖于以前的表现下, 独立增量假设就比较充分。

如果一个计数过程在任一时间区间内发生的事件个数的分布仅取决于时间长度, 则称它具有平稳增量性。换言之, 如果对任意的 $s>0, t_1<t_2$, 在时间区间 $(t_1+s, t_2+s]$ 内发生的事件次数 $N(t_2+s)-N(t_1+s)$ 与在时间区间 $(t_1,t_2]$ 内发生的事件次数 $N(t_2)-N(t_1)$ 有相同的分布, 那么称计数过程具有平稳增量性。

在例 1 中, 当人们不会在某些时段内更倾向于进入商店时, 平稳增量性假设才是合理的。因此, 如果每天都存在一个购物的高峰时段, 那么平稳增量的假设是不成立的。在例 2 中, 只有当人们认为地球上的人口总数基本上是一个常数时(这是大多数人都不支持的观点), 平稳增量假设才可能成立。在例 3 中, 对一场比赛来说, 平稳增量假设是可以接受的。

2. Poisson 过程

Poisson 过程是非常特殊的计数过程, 它也是计数过程中最重要的类型之一, 其定义如下。

定义 1 如果计数过程 $\{N(t), t\geqslant 0\}$ 满足以下条件:

(i) $N(0)=0$。

(ii) N 具有独立增量性。

(iii) 在长度为 t 的时间内事件发生的次数变量服从 Poisson 分布, 参数为 λt, 即对任意的 $s, t\geqslant 0$,

$$\mathbb{P}\{N(t+s)-N(s)=n\}=\mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n=0,1,\cdots$$

那么称 N 是强度参数为 λ $(\lambda>0)$ 的 Poisson 过程。

注意到从定义的条件(iii) 可知 Poisson 过程有平稳增量且 $\mathbb{E}[N(t)]=\lambda t$。参数 λ 称为过程 N 的强度。

为确定一个计数过程是否为 Poisson 过程, 须验证条件(i) 、(ii) 和(iii) 。条件(i) 说明事件的计数是从 0 开始的。条件(ii) 通常可以从对过程的了解情况直接验证。但直接验证条件(iii) 比较困难。为此, 下面给出另一个等价的定义。

首先, 对函数 $f(h)$, 如果 $\lim\limits_{h\downarrow 0}\dfrac{f(h)}{h}=0$, 那么当 $h\downarrow 0$ 时, 称函数 $f(h)$ 为关于 h 的高阶无穷小, 记为 $f(h)\sim o(h)$。

定义 2 如果计数过程 $\{N(t), t \geqslant 0\}$ 满足以下条件:

(i) $N(0)=0$。

(ii) N 具有平稳独立增量性。

(iii) $\mathbb{P}\{N(h)=1\}=\lambda h+o(h)$。

(iv) $\mathbb{P}\{N(h)\geqslant 2\}=o(h)$。

那么称 N 是强度参数为 λ 的 Poisson 过程。

定理1 定义 1 和定义 2 是等价的。

证明 下面仅证明由定义 2 可推出定义 1, 至于如何由定义 1 推出定义 2 这里不再给出。

首先, 对给定的 $u \geqslant 0$, 引入

$$g(t)=\mathbb{E}[\exp\{-uN(t)\}]$$

下面推导关于 $g(t)$ 的微分方程。由平稳独立增量性,

$$\begin{aligned} g(t+h) &= \mathbb{E}[\exp\{-uN(t+h)\}] \\ &= \mathbb{E}[\exp\{-uN(t)\}\exp\{-u[N(t+h)-N(t)]\}] \\ &= \mathbb{E}[\exp\{-uN(t)\}]\mathbb{E}[\exp\{-u[N(t+h)-N(t)]\}] \\ &= g(t)g(h) \end{aligned} \tag{1}$$

由定义 2 的 (iii) 和(iv), 可得

$$\mathbb{P}\{N(h)=0\}=1-\lambda h+o(h)$$

因此, 分别考虑 $N(h)=0, N(h)=1$ 和 $N(h)\geqslant 2$, 并由全概率公式, 可得

$$\begin{aligned} \mathbb{E}[\exp\{-uN(h)\}] &= 1-\lambda h+o(h)+\mathrm{e}^{-u}(\lambda h+o(h))+o(h) \\ &= 1-\lambda h+\mathrm{e}^{-u}\lambda h+o(h) \end{aligned} \tag{2}$$

因此, 由(1)式和(2)式, 可得

$$g(t+h)=g(t)(1-\lambda h+\mathrm{e}^{-u}\lambda h)+o(h)$$

变形即得

$$\frac{g(t+h)-g(t)}{h}=g(t)\lambda(\mathrm{e}^{-u}-1)+\frac{o(h)}{h}$$

令 $h\downarrow 0$, 取极限得到

$$g'(t)=g(t)\lambda(\mathrm{e}^{-u}-1)$$

上式就是

$$(\ln g(t))' = \frac{g'(t)}{g(t)} = \lambda(\mathrm{e}^{-u} - 1)$$

两边取积分，注意到由定义 2 的(i)，可得 $g(0) = 1$。最后得到

$$g(t) = \exp[\lambda t(\mathrm{e}^{-u} - 1)]$$

上述结论表明，$N(t)$ 的 Laplace 变换是 $\mathrm{e}^{\lambda t(\mathrm{e}^{-u}-1)}$，它是均值为 λt 的 Poisson 随机变量的 Laplace 变换。至此即证定义 1 中的(iii)。

关于 Poisson 过程，最后指出以下两点：

(i) 在一定条件下，Poisson 分布是二项分布的近似，这可以解释为什么 $N(t)$ 有 Poisson 分布。具体地讲，把区间 $[0,t]$ 划分为足够多的 k 个相等的部分。由定义 2 的条件(iv)可知，当 k 很大时，对于 k 个子区间的任意一个区间来说，在其中发生两个或两个以上事件的概率趋于 0，即在每个区间内事件发生的次数近似于 Bernoulli 分布变量。因此，由平稳和独立增量性，$N(t)$ 近似服从参数为 (k,p) 的二项分布，其中 $p = \lambda(t/k) + o(t/k)$。当 k 趋近于∞时，可知 $N(t)$ 服从 Poisson 分布，其均值为

$$\lim_{k\to\infty} k\left[\lambda\frac{t}{k} + o\left(\frac{t}{k}\right)\right] = \lambda t$$

(ii) 由定理的证明可知，定义 2 中的条件(ii) (iii) 和(iv)可以改为：对任意的 t，在区间 $(t, t+h)$ 内只发生一次事件的概率是 $\lambda h + o(h)$，而发生两次及两次以上的概率是 $o(h)$。

§2 Poisson 过程的性质

1. 到达时间间隔

对强度参数为 λ 的 Poisson 过程 N，记第一次事件发生的时间为 T_1。进一步，当 $n \geqslant 2$ 时，以 T_n 表示第 $n-1$ 次事件发生到第 n 次事件发生之间的时间间隔。称 $\{T_n, n = 1, 2, \cdots\}$ 为到达时间间隔序列。例如，设 $T_1 = 5, T_2 = 10$ 那么 Poisson 过程的第一次事件发生在时刻 5，第二次事件发生在时刻 15。

下面确定 T_n 的分布。首先，注意到事件 $\{T_1 > t\}$ 发生当且仅当 Poisson 过程在区间 $[0,t]$ 中没有事件发生，因此

$$\mathbb{P}\{T_1 > t\} = \mathbb{P}\{N(t) = 0\} = \mathrm{e}^{-\lambda t}$$

即 T_1 服从均值为 $1/\lambda$ 的指数分布。进一步，

$$\mathbb{P}\{T_2 > t\} = \mathbb{E}[\mathbb{P}\{T_2 > t | T_1\}]$$

由平稳独立增量性,

$$\begin{aligned}\mathbb{P}\{T_2 > t|T_1 = s\} &= \mathbb{P}\{(s, s+t]\text{内没有事件发生}|T_1 = s\}\\&= \mathbb{P}\{(s, s+t]\text{内没有事件发生}\}\\&= \mathrm{e}^{-\lambda t}\end{aligned} \tag{3}$$

从而 T_2 也服从均值为 $1/\lambda$ 的指数分布, 而且 T_2 与 T_1 独立。重复以上论述, 即得

性质 1 $\{T_n, n = 1, 2, \cdots\}$ 是独立同分布的指数随机变量, 均值为 $1/\lambda$。

性质 1 的结论很直观。由平稳独立增量的假设, 在任意时刻, Poisson 过程在概率意义下重新开始, 在某一时刻及其以后发生的事件与该时刻以前发生的事件是独立的。换言之, Poisson 过程具有无记忆性。由概率论中的结论, 具有无记忆性的连续型分布只有指数分布。

2. 等待时间

记 S_n 为第 n 次事件的到达时间,

$$S_n = \sum_{i=1}^{n} T_i, \quad n \geqslant 1$$

S_n 也称为第 n 次事件发生的等待时间。

由性质 1 可知, S_n 服从参数为 (n, λ) 的 Gamma 分布, 即 S_n 的概率密度函数为

$$f_{S_n}(t) = \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t \geqslant 0 \tag{4}$$

(4)式也可通过另外一种方式得到。注意到第 n 次事件在时刻 t 之前发生, 等价于到时刻 t 时至少发生了 n 次事件, 即 $\{S_n \leqslant t\}$ 和 $\{N(t) \geqslant n\}$ 是等价的。因此,

$$F_{S_n}(t) = \mathbb{P}\{S_n \leqslant t\} = \mathbb{P}\{N(t) \geqslant n\} = \sum_{j=n}^{\infty} \mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!}$$

求导数可得

$$\begin{aligned}f_{S_n}(t) &= -\sum_{j=n}^{\infty} \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!} + \sum_{j=n}^{\infty} \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-1}}{(j-1)!}\\&= \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!} + \sum_{j=n+1}^{\infty} \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^j}{j!}\\&= \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}\end{aligned}$$

另外, 还可以用下面的方法来得 S_n 的密度函数,

$$\begin{aligned}\mathbb{P}\{t < S_n < t+h\} &= \mathbb{P}\{N(t)=n-1, \text{而且}(t,t+h)\text{内发生一次事件}\} + o(h) \\ &= \mathbb{P}\{N(t)=n-1\}\mathbb{P}\{(t,t+h)\text{内发生一次事件}\} + o(h) \\ &= \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}[\lambda h + o(h)] + o(h) \\ &= \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}h + o(h)\end{aligned}$$

在上式两边同除以 h, 令 $h \downarrow 0$, 可得

$$f_{S_n}(t) = \lambda \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}$$

例 4 假设人们移居到某一地区服从 $\lambda = 1$ 人/天的 Poisson 过程, 那么

(i) 第 10 个移民到达的时间期望为

$$\mathbb{E}[S_{10}] = 10\frac{1}{\lambda} = 10(\text{天})$$

(ii) 第 10 个和第 11 个移民到来的时间间隔超过两天的概率

$$\mathbb{P}\{T_{11} > 2\} = \mathrm{e}^{-2\lambda} = \mathrm{e}^{-2} \approx 0.133$$

从性质 1 出发, 可以得到 Poisson 过程的另一种定义方法。首先给出均值为 $1/\lambda$ 的独立同分布指数分布随机变量序列 $\{T_n, n \geqslant 1\}$, 然后定义一个计数过程 N, 使得第 n 次事件的发生时间为

$$S_n \equiv T_1 + T_2 + \cdots + T_n$$

那么 $\{N(t), t \geqslant 0\}$ 就是强度为 λ 的 Poisson 过程。此时 $N(t)$ 的具体形式为

$$N(t) = \begin{cases} 0, & 0 \leqslant t < T_1 \\ 1, & T_1 \leqslant t < S_2 \\ \vdots, & \vdots \end{cases}$$

3. Poisson 过程的分解

对强度参数为 λ 的 Poisson 过程 $\{N(t), t \geqslant 0\}$, 如果按某一属性, 可把事件归为两类: 第 I 类或第 II 类。假设每一次事件可归为第 I 类的概率是 p, 可归为第 II 类的概率为 $1-p$, 而且独立于其他事件。例如, 假设到达某商店的顾客人数构成 Poisson 过程, 而且每一个顾客是男性或女性的概率各为1/2,那么第 I 类事件对应于男性顾客的到达, 第 II 类事件对应于女性顾客的到达。

记 $N_1(t)$ 和 $N_2(t)$ 分别表示在区间 $[0,t]$ 内发生的第 I 类事件的次数和第 II 类事件的次数。注意到 $N(t) = N_1(t) + N_2(t)$。

性质 2 $\{N_1(t), t \geqslant 0\}$ 和 $\{N_2(t), t \geqslant 0\}$ 都是 Poisson 过程, 相应的参数分别为 λp 和 $\lambda(1-p)$, 而且这两个过程相互独立。

证明 首先计算联合概率 $\mathbb{P}\{N_1(t)=n, N_2(t)=m\}$。由全概率公式可得

$$
\begin{aligned}
&\mathbb{P}\ \{\ N_1(t)=n, N_2(t)=m\} \\
&\quad = \sum_{k=0}^{\infty} \mathbb{P}\{N_1(t)=n, N_2(t)=m|N(t)=k\}\mathbb{P}\{N(t)=k\}
\end{aligned}
$$

一方面, 为使区间 $[0,t]$ 内有 n 个第 I 类事件和 m 个第 II 类事件发生, 那么须有 $n+m$ 个事件发生。从而当 $k \neq n+m$ 时,

$$
\mathbb{P}\{N_1(t)=n, N_2(t)=m|N(t)=k\}=0
$$

因此

$$
\begin{aligned}
&\mathbb{P}\ \{\ N_1(t)=n, N_2(t)=m\} \\
&\quad = \mathbb{P}\{N_1(t)=n, N_2(t)=m|N(t)=n+m\}\mathbb{P}\{N(t)=n+m\} \\
&\quad = \mathbb{P}\{N_1(t)=n, N_2(t)=m|N(t)=n+m\}\mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n+m}}{(n+m)!}
\end{aligned}
$$

另一方面, 在 $n+m$ 个事件中, 有 n 个第 I 类事件和 m 个第 II 类事件的概率为

$$
\begin{pmatrix} n+m \\ n \end{pmatrix} p^n(1-p)^m
$$

从而

$$
\begin{aligned}
\mathbb{P}\{N_1(t)=n, N_2(t)=m\} &= \begin{pmatrix} n+m \\ n \end{pmatrix} p^n(1-p)^m \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{n+m}}{(n+m)!} \\
&= \mathrm{e}^{-\lambda tp}\frac{(\lambda tp)^n}{n!}\mathrm{e}^{-\lambda t(1-p)}\frac{(\lambda t(1-p))^m}{m!}
\end{aligned} \tag{5}
$$

(5)式表明, $N_1(t)$ 和 $N_2(t)$ 都有 Poisson 分布, 参数分别为 λpt 和 $\lambda(1-p)t$, 而且 $N_1(t)$ 和 $N_2(t)$ 是独立的。

其次, 需要再验证定义 1 的其他条件成立, 从而证明 $\{N_1(t), t \geqslant 0\}$ 是强度参数为 λp 的 Poisson 过程, $\{N_2(t), t \geqslant 0\}$ 是强度参数为 $\lambda(1-p)$ 的 Poisson 过程。此处不再展开。

$\{N_1(t), t \geqslant 0\}$ 和 $\{N_2(t), t \geqslant 0\}$ 都是 Poisson 过程比较容易理解, 但令人惊奇的是它们的独立性。假设到达某银行的顾客人数服从 $\lambda=1$ 人/小时的 Poisson

过程, 而且每个顾客是男性或女性的概率各为 1/2。设想在最初的 10 个小时内有 100 位男顾客到达银行, 那么在这段时间内女顾客到达人数的期望是多少呢? 有人也许会想当然地认为: 因为男顾客到达的人数是 100, 而且每个顾客是男性的概率是 1/2, 所以女顾客到达人数的期望也应该是 100。但是, 由性质 2 可知, 这样的推理并不正确。在最初的 10 小时内女顾客到达人数的期望是 5, 它与这段时间内的男顾客到达人数是独立的。

为了从直观上理解性质 2, 可作如下推理: 把区间 $[0,t]$ 划分为长度均为 t/n 的 n 个子区间, 这里 n 是个很大的正整数, 那么每个子区间包含单个事件的概率近似为 $\lambda t/n$。因为每个事件是第 I 类的概率为 p, 因此在 n 个子区间的任意一个小区间内, 或者没有事件发生, 或者发生第 I 类事件, 或者发生第 II 类事件, 相应的概率分别为

$$1-\frac{\lambda t}{n},\quad \frac{\lambda t}{n}p,\quad \frac{\lambda t}{n}(1-p)$$

由此, 从二项分布的极限分布是 Poisson 分布这一结论出发, 可以得出 $N_1(t)$ 和 $N_2(t)$ 分别服从均值为 λtp 和 $\lambda t(1-p)$ 的 Poisson 分布。

例 5 如果迁居到 A 地的移民服从 $\lambda=10$ 人/周的 Poisson 过程, 同时假如每个移民是英国后裔的概率是 1/12, 确定在二月份没有英国后裔移民迁居到 A 地区的概率。

此时, $t=4$, $p=1/12$, 由性质 2 可知, 在二月份英国后裔移民到 A 地的数目服从均值为 $10\times 4\times(1/12)=10/3$ 的 Poisson 分布。因此在二月份没有英国后裔移民迁居到 A 地区的概率为 $\mathrm{e}^{-10/3}=0.0357$。

性质 2 的结论很容易推广到有 r 个子类的情况。下面给出两个例子。

例 6 (员工流动问题) 考察一个系统, 其中的个体是相互独立的, 每个个体随时都能转移到 r 个可能状态中, 每个个体状态的改变过程构成转移概率为 $P_{ij}, i,j=1,\cdots,r$ 的 Markov 链, 即在某时刻处于状态 i 的个体, 在下一时刻以概率 P_{ij} 转移到状态 j。另外, 每个个体在系统中的状态转移都是相互独立的。设在最初处于状态 $1,2,\cdots,r$ 的个体数都是独立的 Poisson 随机变量, 均值分别为 $\lambda_1,\lambda_2,\cdots,\lambda_r$。确定在时刻 n 处于状态 $1,2,\cdots,r$ 的个体数的联合分布。

设状态的改变过程构成一个 Poisson 过程。对于给定的 i, 引入 $N_j(i)$, $j=1,\cdots,r$, 它表示最初处于状态 i 而在时刻 n 处于状态 j 的个体数。每一个最初处于状态 i 的个体在时刻 n 处于状态 j 的概率为 $P_{ij}^{(n)}$, 这里 $P_{ij}^{(n)}$ 表示转移概率为 P_{ij} 的 Markov 链的 n 步转移概率。因此 $\{N_j(i), j=1,\cdots,r\}$ 是相互独立的均值为 $\{\lambda_i P_{ij}^{(n)}, j=1,\cdots,r\}$ 的 Poisson 随机变量。因为相互独立的 Poisson 随机变量之和仍然是 Poisson 随机变量, 由此可得在时刻 n 处于状态 j 的个体数 $\sum\limits_{i=1}^{r} N_j(i)$

是均值为 $\sum_i \lambda_i P_{ij}^{(n)}$ 的 Poisson 随机变量。

例 7 (票券收集问题) 有 m 种不同类型的票券。某人每次收集到的票券都独立于他以前收集到的票券, 且收集到第 j 种票券的概率是 p_j, $\sum_{j=1}^{m} p_j = 1$。N 表示为凑齐完整的票券类型(使得每一种类型票券至少都有一张)所需要收集的票券数, 求 $\mathbb{E}[N]$。

以 N_j 表示为了收集到第 j 种票券所必需收集的票券数, 那么 N 可表示为

$$N = \max_{1 \leqslant j \leqslant m} N_j$$

虽然每一个 N_j 都服从参数为 p_j 的几何分布, 但随机变量 N_j 不相互独立, 因此上面给出 N 的表示意义不大。

下面把问题转化为独立随机变量的最大值问题。为此, 设收集票券的过程构成 $\lambda = 1$ 的 Poisson 过程。在某一时刻收集到了类型为 j 的票券, 对应于 Poisson 过程的第 j 种类型事件发生, $1 \leqslant j \leqslant m$。以 $N_j(t)$ 表示到时刻 t 时收集到的第 j 种票券数, 由性质 2 可知 $\{N_j, j = 1, \cdots, m\}$ 是相互独立的 Poisson 过程, 均值分别为 $\lambda p_j = p_j, j = 1, \cdots, m$。以 X_j 表示第 j 个过程首次事件发生的时刻, 引入

$$X = \max_{1 \leqslant j \leqslant m} X_j$$

X 表示为凑齐完整的收集所需要的时间。因为 X_j 服从均值为 p_j 的指数分布, 而且 $\{X_j, j = 1, \cdots, m\}$ 相互独立, 由此可得

$$\mathbb{P}\{X < t\} = \mathbb{P}\{\max X_j < t\} = \mathbb{P}\{X_j < t, j = 1, \cdots, m\} = \prod_{j=1}^{m} (1 - \mathrm{e}^{-p_j t})$$

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\} \mathrm{d}t = \int_0^\infty \left\{ 1 - \prod_{j=1}^{m} (1 - \mathrm{e}^{-p_j t}) \right\} \mathrm{d}t \tag{6}$$

下面把收集到一套完整票券所需要的时间的期望值 $\mathbb{E}[X]$ 与所需收集的票券数目 $\mathbb{E}[N]$ 联系起来。为此, 以 T_i 表示收集票券的 Poisson 过程的第 i 个事件的到达时间间隔。显然, $X = \sum_{i=1}^{N} T_i$。因为 T_i 服从参数为 1 的指数分布且相互独立, 所以 $\mathbb{E}[X] = \mathbb{E}[N]$, 即 $\mathbb{E}[N]$ 由(6)式给出。

4. 一个概率计算问题

下面考虑关于 Poisson 过程的一个概率计算问题。设 $\{N_1(t), t \geqslant 0\}$ 和 $\{N_2(t), t \geqslant 0\}$ 是相互独立的 Poisson 过程, 强度参数分别为 λ_1 和 λ_2。以 S_n^1 表示 N_1 的第 n

次事件发生的时刻,S_m^2 表示 N_2 的第 m 次事件发生的时刻。现在计算如下概率

$$\mathbb{P}\left\{S_n^1 < S_m^2\right\}$$

先考虑 $n = m = 1$ 的特殊情形。由性质 1, $N_1(t)$ 首次事件发生时刻 S_1^1 和 $N_2(t)$ 首次事件发生时刻 S_1^2 都是指数随机变量, 均值分别为 $1/\lambda_1$ 和 $1/\lambda_2$, 从而

$$\mathbb{P}\left\{S_1^1 < S_1^2\right\} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{7}$$

(7)式可由概率论的初等方法计算得到。

现在考虑 $N_1(t)$ 的两个事件发生在 $N_2(t)$ 的一个事件前的概率, 即 $\mathbb{P}\left\{S_2^1 < S_1^2\right\}$。引用如下推理：为使 $N_1(t)$ 的两个事件发生在 $N_2(t)$ 的一个事件前, 首先, 最初发生的事件须是 $N_1(t)$ 的事件, (7)式给出了这个概率。另外, 给定最初发生的事件来自 $N_1(t)$, 要使 $S_2^1 < S_1^2$, 第二个事件须来自 $N_1(t)$, 同时, 由 Poisson 过程的无记忆性, 在首次事件发生的条件下, 两个过程又都重新开始, 因此, 条件概率也是 $\lambda_1/(\lambda_1 + \lambda_2)$。从而

$$\mathbb{P}\left\{S_2^1 < S_1^2\right\} = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2$$

由上面的推理可知, 发生的每次事件来自 $N_1(t)$ 的条件概率是 $\lambda_1/(\lambda_1 + \lambda_2)$, 来自 $N_2(t)$ 的概率是 $\lambda_2/(\lambda_1 + \lambda_2)$, 而且独立于所有以前发生的事件。换言之, $N_1(t)$ 的 n 个事件发生在 $N_2(t)$ 的 m 个事件发生之前, 等价于在投币试验中, n 次正面出现在 m 次反面之前, 其中出现正面的概率为 $\lambda_1/(\lambda_1 + \lambda_2)$。这样的事件要发生, 当且仅当前 $n + m - 1$ 次投币试验中至少出现 n 次正面。最后, 所求的概率为

$$\mathbb{P}\left\{S_n^1 < S_m^2\right\} = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n+m-1-k}$$

5. 到达时间的条件分布

设在时刻 t 之前, Poisson 过程的一个事件已经发生, 如何确定事件发生时间的分布？由 Poisson 过程的平稳独立增量性, 直观上有充分的理由认为, 事件发生的时间应服从区间 $[0, t]$ 上的均匀分布。这一结论是正确的, 而且容易验证如下：对 $s \leqslant t$,

$$\begin{aligned}
\mathbb{P}\{T_1 < s | N(t) = 1\} &= \frac{\mathbb{P}\{T_1 < s, N(t) = 1\}}{\mathbb{P}\{N(t) = 1\}} \\
&= \frac{\mathbb{P}\{[0, s)\text{内发生一次事件}, [s, t]\text{内没有事件发生}\}}{\mathbb{P}\{N(t) = 1\}} \\
&= \frac{\mathbb{P}\{[0, s)\text{内发生一次事件}\}\mathbb{P}\{[s, t]\text{内没有事件发生}\}}{\mathbb{P}\{N(t) = 1\}} \\
&= \frac{\lambda s \mathrm{e}^{-\lambda s} \mathrm{e}^{-\lambda(t-s)}}{\lambda t \mathrm{e}^{-\lambda t}} = \frac{s}{t}
\end{aligned}$$

上述结论可以进一步推广，为此先简单介绍顺序统计量的概念。

给定 n 个随机变量 $Y_1, Y_2, \cdots, Y_n$, 记 $Y_{(k)}$ 是 $Y_1, \cdots, Y_n$ 中第 k 个最大值, 那么就称 $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ 是对应于 $Y_1, Y_2, \cdots, Y_n$ 的顺序统计量。例如, 设 $n=3, Y_1=4, Y_2=5, Y_3=1$, 那么 $Y_{(1)}=1, Y_{(2)}=4, Y_{(3)}=5$。设 $\{Y_i, i=1,\cdots,n\}$ 是独立同分布连续型随机变量, 概率密度函数为 f, 那么顺序统计量 $\{Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}\}$ 的联合密度函数为

$$f(y_1, y_2, \cdots, y_n) = n!\prod_{i=1}^{n} f(y_i),\ y_1 < y_2 < \cdots < y_n$$

为说明上式成立, 注意到

(i) 当 $(Y_1, Y_2, \cdots, Y_n)$ 分别取值于 $(y_1, y_2, \cdots, y_n)$ 的 $n!$ 个排列时, $(Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}) = (y_1, y_2, \cdots, y_n)$。

(ii) 对 $1, 2, \cdots, n$ 的排列 $i_1, \cdots, i_n$, $(Y_1, Y_2, \cdots, Y_n) = (y_{i_1}, \cdots, y_{i_n})$ 的概率密度是 $\prod_{j=1}^{n} f(y_{i_j}) = \prod_{j=1}^{n} f(y_i)$。

如果 $\{Y_i, i=1,\cdots,n\}$ 是 $[0,t]$ 上独立的均匀分布随机变量, 那么由前面的结论可得到顺序统计量 $\{Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}\}$ 的联合密度函数为

$$f(y_1, y_2, \cdots, y_n) = n!\frac{1}{t^n},\ 0 < y_1 < y_2 < \cdots < y_n < t$$

定理 2 在 $N(t)=n$ 的条件下, n 个到达时间 $S_1, \cdots, S_n$ 的联合密度等于 $[0,t]$ 上 n 个独立的均匀分布随机变量的顺序统计量的密度。

证明 为得到给定条件 $N(t)=n$ 下 $S_1, \cdots, S_n$ 的条件密度, 在 $0 < s_1 < \cdots < s_n < t$ 时, 事件 $\{S_1 = s_1, S_2 = s_2, \cdots, S_n = s_n, N(t) = n\}$ 等价于前 $n+1$ 次到达时间间隔满足

$$\{T_1 = s_1, T_2 = s_2 - s_1, \cdots, T_n = s_n - s_{n-1}, T_{n+1} > t - s_n\}$$

因此, 由性质 1 可以得到如下的在给定 $N(t)=n$ 情况下 $S_1, \cdots, S_n$ 的条件联合密度

$$\begin{aligned} f(s_1, \cdots, s_n | n) &= \frac{f(s_1, \cdots, s_n, n)}{\mathbb{P}\{N(t) = n\}} \\ &= \frac{\lambda e^{-\lambda s_1} \lambda e^{-\lambda(s_2 - s_1)} \cdots \lambda e^{-\lambda(s_n - s_{n-1})} e^{-\lambda(t - s_n)}}{e^{-\lambda t}(\lambda t)^n / n!} \\ &= \frac{n!}{t^n} \end{aligned}$$

定理 2 也可以表述为, 在 $[0,t]$ 内发生 n 次事件的条件下, 事件的发生时间是 $[0,t]$ 上 n 个相互独立的均匀分布随机变量。

§3 Poisson 过程的应用举例

本节通过一些例子, 说明§2节的定理 2 的各种应用。

由 § 2 节的性质 2, 如果 Poisson 过程的每一次事件归为第 I 类的概率为 p, 归为第 II 类的概率为 $1-p$, 那么由第 I 类和第 II 类事件分别构成的计数过程是相互独立的 Poisson 过程, 强度参数分别为 λp 和 $\lambda(1-p)$。

进一步, 现在设想有 k 个事件类, 而且一个事件被归为第 i 类的概率取决于该事件发生的时间。具体地, 设发生在时刻 y 的事件被归为第 i 类的概率为 $P_i(y)$, $\sum\limits_{i=1}^{k} P_i(y)=1$, 同时该事件与以前发生的事件独立。由§2节的定理 2 可以证明下面性质。

性质 3 以 $N_i(t)$ 表示到时刻 t 时第 i 类事件的个数, 那么 $\{N_i, i=1,\cdots,k\}$ 是相互独立的 Poisson 过程, 强度参数为

$$\mathbb{E}[N_i(t)]=\lambda\int_0^t P_i(s)\mathrm{d}s$$

证明 为证明命题, 可求联合概率 $\mathbb{P}\{N_i(t)=n_i, i=1,\cdots,k\}$。首先, 注意到总的事件个数为 $\sum\limits_{i=1}^{k} n_i$。因此,

$$\begin{aligned}&\mathbb{P}\{N_1(t)=n_1,\cdots,N_k(t)=n_k\}\\&=\mathbb{P}\left\{N_1(t)=n_1,\cdots,N_k(t)=n_k\middle|N(t)=\sum_{i=1}^{k}n_i\right\}\mathbb{P}\left\{N(t)=\sum_{i=1}^{k}n_i\right\}\end{aligned}$$

考虑在 $[0,t]$ 内发生的任意事件, 如果它在时刻 s 发生, 那么它是第 i 类事件的概率是 $P_i(s)$。另外, 由§2节的定理 2, 事件发生的时间在 $(0,t)$ 内均匀分布, 因此, 在 $(0,t)$ 内发生的事件为第 i 类事件的概率为

$$P_i=\frac{1}{t}\int_0^t P_i(s)\mathrm{d}s$$

考虑条件概率

$$\mathbb{P}\left\{N_1(t)=n_1,\cdots,N_k(t)=n_k\middle|N(t)=\sum_{i=1}^{k}n_i\right\}$$

它是多项分布的概率, $\sum\limits_{i=1}^{k} n_i$ 中的每次试验以概率 P_i 出现第 i 种类型事件,

$$\mathbb{P}\left\{N_1(t)=n_1,\cdots,N_k(t)=n_k\middle|N(t)=\sum_{i=1}^{k}n_i\right\}=\frac{\left(\sum\limits_{i}n_i\right)!}{n_1!\cdots n_k!}P_1^{n_1}\cdots P_k^{n_k}$$

最后,

$$\mathbb{P}\{N_1(t)=n_1,\cdots,N_k(t)=n_k\}=\frac{\left(\sum_i n_i\right)!}{n_1!\cdots n_k!}P_1^{n_1}\cdots P_k^{n_k}\mathrm{e}^{-\lambda t}\frac{(\lambda t)^{\sum_i n_i}}{\left(\sum_i n_i\right)!}$$

$$=\prod_{i=1}^{k}\mathrm{e}^{-\lambda P_i t}\frac{(\lambda P_i t)^{n_i}}{n_i!}$$

下面举例说明性质 3 的应用。

例 8 (无限服务台队列) 假设顾客到达某服务台构成参数为 λ 的 Poisson 过程。每当顾客到达服务台即有一名服务员为其服务, 设有充分多的服务员, 顾客到达服务台后立即接受服务而无需等待。设每个顾客到达后的服务时间服从某分布 G。以 $X(t)$ 和 $Y(t)$ 分别表示到时刻 t 为止已经接受和正在接受服务的顾客人数, 现在考虑这一随机服务系统的效率, 也就是要计算 $X(t)$ 和 $Y(t)$ 的分布。

为此, 首先把顾客归为两类: 到时刻 t 为止已经接受服务的归为第 I 类, 而到时刻 t 为止仍正在接受服务的归为第 II 类。假设某顾客在时刻 s $(s\leqslant t)$ 进入服务台, 如果服务时间少于 $t-s$, 那么该顾客归为第 I 类。由于服务时间的分布为 G, 可知该事件发生的概率为 $G(t-s)$。类似地, 如果顾客在时刻 s $(s\leqslant t)$ 进入服务台, 那么该顾客归为第 II 类的概率为 $\bar{G}(t-s)=1-G(t-s)$。由性质 3 可知, 到时刻 t 为止已接受服务的顾客人数 $X(t)$ 服从 Poisson 分布, 均值为

$$\mathbb{E}[X(t)]=\lambda\int_0^t G(t-s)\mathrm{d}s=\lambda\int_0^t G(y)\mathrm{d}y$$

而在时刻 t 正在接受服务的顾客人数 $Y(t)$ 也服从 Poisson 分布, 均值为

$$\mathbb{E}[Y(t)]=\lambda\int_0^t \bar{G}(t-s)\mathrm{d}s=\lambda\int_0^t \bar{G}(y)\mathrm{d}y$$

进一步, $X(t)$ 与 $Y(t)$ 相互独立。

现在计算 $Y(t)$ 与 $Y(t+s)$ 的联合概率。为此, 对顾客进行以下分类:

(1) 在时刻 t 之前到达并在 t 与 $t+s$ 之间完成服务的顾客。

(2) 在时刻 t 之前到达并在 $t+s$ 之后完成服务的顾客。

(3) 在时刻 t 与 $t+s$ 之间到达并在 $t+s$ 之后完成服务的顾客。

(4) 其他。

因此在时刻 y 到达的顾客归为第 i 类的概率是

$$P_1(y)=\begin{cases}G(t+s-y)-G(t-y), & y<t\\ 0, & \text{其他}\end{cases}$$

$$P_2(y) = \begin{cases} \bar{G}(t+s-y), & y < t \\ 0, & \text{其他} \end{cases}$$

$$P_3(y) = \begin{cases} \bar{G}(t+s-y), & t < y < t+s \\ 0, & \text{其他} \end{cases}$$

$$P_4(y) = 1 - P_1(y) - P_2(y) - P_3(y)$$

以 $N_i = N_i(s+t), i = 1,2,3$ 表示第 i 类事件发生的次数, 由性质 3 可知, $N_i, i = 1,2,3$ 是独立的 Poisson 随机变量, 均值分别为

$$\mathbb{E}[N_i] = \lambda \int_0^{t+s} P_i(y)\mathrm{d}y, \quad i = 1,2,3$$

注意到 $Y(t) = N_1 + N_2, Y(t+s) = N_2 + N_3, Y(t)$ 与 $Y(t+s)$ 的联合概率为

$$\begin{aligned} \mathbb{P}\{Y(t) = i, Y(t+s) = j\} &= \mathbb{P}\{N_1 + N_2 = i, N_2 + N_3 = j\} \\ &= \sum_{l=0}^{\min(i,j)} \mathbb{P}\{N_2 = l, N_1 = i-l, N_3 = j-l\} \\ &= \sum_{l=0}^{\min(i,j)} \mathbb{P}\{N_2 = l\}\mathbb{P}\{N_1 = i-l\}\mathbb{P}\{N_3 = j-l\} \end{aligned}$$

另外,

$$\begin{aligned} \mathrm{Cov}[Y(t), Y(t+s)] &= \mathrm{Cov}(N_1 + N_2, N_2 + N_3) \\ &= \mathrm{Cov}(N_2, N_2) = \mathrm{Var}(N_2) \\ &= \lambda \int_0^t \bar{G}(t+s-y)\mathrm{d}y \\ &= \lambda \int_0^t \bar{G}(u+s)\mathrm{d}u \end{aligned}$$

在上面的推导中, 用到了 $N_i, i = 1,2,3$ 的相互独立性, 以及 Poisson 随机变量的方差与期望相等的结论。

例 9 (相遇次数最小化) 设想进入某单行道高速公路的汽车构成参数为 λ 的 Poisson 过程。汽车由位置 a 点进入, 由位置 b 点离开。每辆汽车的速度是个常数, 不同汽车的速度随机地由分布 G 来决定。假设在时刻 s 驶入高速公路并且可以自由选择车速, 那么以多大的速度行驶能使得与其他车辆相遇次数的期望值最小? 当超过别的车或者被别的车超过时, 就称为一次相遇。

下面要证明对于很大的 s 来说, 能使得与其他车辆相遇次数的期望值最小的速度是速度分布 G 的中位数。设所选择的速度是 x, 记 $d = b - a$ 表示高速公路的

长度。选定了速度 x 后, 那么汽车将在时刻 s 进入公路, 而在时刻 $s+t_0$ 离开公路, 这里 $t_0=d/x$ 为行驶时间。

汽车进入高速公路构成参数为 λ 的 Poisson 过程。每辆汽车都根据分布 G 来选择一个速度 X, 从而行驶时间 $T=d/X$。记 F 表示行驶时间 T 的分布,

$$F(t)=\mathbb{P}\{T\leqslant t\}=\mathbb{P}\{d/X\leqslant t\}=\mathbb{P}\{X\geqslant d/t\}=\bar{G}(d/t)$$

如果另外一辆汽车在时刻 t 进入了高速公路, 就称发生了一次事件, 而且如果这辆车与开始参照的车相遇, 就把事件归为第 I 类。开始参照的车将在时刻 s 进入公路并在时刻 $s+t_0$ 离开公路。因此, 参照车将与在时刻 s 之前进入而在时刻 $s+t_0$ 之后离开公路的车相遇(参照车在公路上超过别人的车), 还将与在时刻 s 后进入公路但在时刻 $s+t_0$ 前离开公路的车相遇(别人的车超过参照车)。总之, 在时刻 t 进入公路的汽车, 如果它的行驶时间 T 满足以下条件, 那么两辆车将会相遇：或者

$$t+T>s+t_0,\quad t<s$$

或者

$$t+T<s+t_0,\quad s<t<s+t_0$$

如前所述, 在时刻 t 发生的事件可归为第 I 类的概率 $P(t)$ 为

$$P(t)=\begin{cases}\mathbb{P}\{t+T>s+t_0\}=\bar{F}(s+t_0-t), & t<s\\ \mathbb{P}\{t+T<s+t_0\}=F(s+t_0-t), & s<t<s+t_0\\ 0, & t>s+t_0\end{cases}$$

由性质 3, 第 I 类事件的总次数服从 Poisson 分布, 其均值为

$$\begin{aligned}\lambda\int_0^\infty P(t)\mathrm{d}t&=\lambda\int_0^s\bar{F}(s+t_0-t)\mathrm{d}t+\lambda\int_s^{s+t_0}F(s+t_0-t)\mathrm{d}t\\&=\lambda\int_{t_0}^{s+t_0}\bar{F}(y)\mathrm{d}y+\lambda\int_0^{t_0}F(y)\mathrm{d}y\end{aligned}$$

为求使上述期望值达到最小的 t_0, 对上式取导数, 得到

$$\frac{\mathrm{d}}{\mathrm{d}t_0}\left\{\lambda\int_0^\infty P(t)\mathrm{d}t\right\}=\lambda\{\bar{F}(s+t_0)-\bar{F}(t_0)+F(t_0)\}$$

令上式右边为 0, 注意到当 s 很大时, $\bar{F}(s+t_0)\approx 0$, 因此最优的行驶时间 t_0 满足

$$F(t_0)-\bar{F}(t_0)=0$$

对上式变形,

$$F(t_0)-[1-F(t_0)]=0$$

$$F(t_0) = 0.5$$

从而最优的行驶时间是行驶时间分布的中位数。因为 $X = d/T$, 对应的最优速度 x_0 是速度分布的中位数, 即

$$G(x_0) = 0.5$$

综上所述, 对于任意的速度 x 来说, 与其他汽车相遇的次数是 Poisson 随机变量, 而且当 x 取速度分布 G 的中位数时相遇次数的均值达到最小。

例 10 (对 HIV 病毒携带者人数的估计) 当一个人感染了 HIV 病毒时, 在疾病症状出现之前会有相当长的一段潜伏期, 这使得公共卫生部门对在某一时期人群中已感染病毒的人数估计十分困难。下面对该现象建立一个近似模型, 以得到对已感染人群的粗略估计。

假设感染 HIV 病毒的人群构成未知参数为 λ 的 Poisson 过程。假设从感染病毒到出现疾病症状的时间是服从已知分布 G 的随机变量, 而且对于不同的个体, 潜伏期是相互独立的。

以 $N_1(t)$ 表示到时刻 t 为止已经表现出疾病症状的人数。另外, 以 $N_2(t)$ 表示 HIV 呈阳性但到时刻 t 为止仍未表现出疾病症状的人数。因为在时刻 s 感染病毒的个体到时刻 t 为止表现出症状的概率为 $G(t-s)$, 没有表现出症状的概率为 $\bar{G}(t-s)$, 所以由性质 3 可得 $N_1(t)$ 和 $N_2(t)$ 是独立的 Poisson 随机变量, 均值分别为

$$\mathbb{E}[N_1(t)] = \lambda \int_0^t G(t-s)\mathrm{d}s = \lambda \int_0^t G(y)\mathrm{d}y$$

$$\mathbb{E}[N_2(t)] = \lambda \int_0^t \bar{G}(t-s)\mathrm{d}s = \lambda \int_0^t \bar{G}(y)\mathrm{d}y$$

如果 λ 是已知的, 那么可以用它来求到时刻 t 为止已感染病毒但尚未有症状表现的 $N_2(t)$ 的估计 $\mathbb{E}[N_2(t)]$。但这里 λ 是未知的, 为此需估计 λ。如果已知 $N_1(t) = n_1$, 那么可以用它对均值 $\mathbb{E}[N_1(t)]$ 进行估计,

$$n_1 \approx \mathbb{E}[N_1(t)] = \lambda \int_0^t G(y)\mathrm{d}y$$

从而 λ 的估计值 $\hat{\lambda}$ 可取为

$$\hat{\lambda} = \frac{n_1}{\displaystyle\int_0^t G(y)\mathrm{d}y}$$

由此可以得到对已经感染了病毒但尚未表现出症状的人数估计,

$$\hat{N}_2(t) = \hat{\lambda} \int_0^t \bar{G}(y)\mathrm{d}y = \frac{n_1 \displaystyle\int_0^t \bar{G}(y)\mathrm{d}y}{\displaystyle\int_0^t G(y)\mathrm{d}y}$$

例如, 假设 G 是均值为 μ 的指数分布, $\bar{G}(y)=\mathrm{e}^{-y/\mu}$ 通过简单计算得到

$$\hat{N}_2(t)=\frac{n_1\mu(1-\mathrm{e}^{-t/\mu})}{t-\mu(1-\mathrm{e}^{-t/\mu})}$$

现举例说明, 如取 $t=16$ 年, $\mu=10$ 年, $n_1=22$ 万人, 那么在时刻 16 时已经感染了病毒但尚未表现出症状的人数估计为

$$\frac{220(1-\mathrm{e}^{-1.6})}{16-10(1-\mathrm{e}^{-1.6})}=21.90\ (\text{万人})$$



第三章 Brown 运动

§1 Brown 运动的定义及一些基本性质

1. 定义

首先考虑一维对称随机游动。假设有一个粒子初始位置为 0, 它在直线上作随机游动, 在每个单位时间内等可能地向左或向右移动一个单位长度, 这是一个 Markov 链, 它的一步转移概率是 $p_{i,i+1} = p_{i,i-1} = 0.5, i = 0, \pm 1, \cdots$。现在改变时间与空间单位, 使得在越来越短的时间间隔里移动的步长越来越小, 以一定的方式取极限, 那么极限过程就是 Brown 运动。

确切地说, 设每隔 Δt 时间粒子等可能地向左或向右移动 Δx 的距离, 如果以 $X(t)$ 表示时刻 t 粒子的位置, 那么

$$X(t) = \Delta x(X_1 + \cdots + X_{[t/\Delta t]}) \tag{1}$$

在(1)式中, $[t/\Delta t]$ 表示 $t/\Delta t$ 的整数部分, $\{X_1, X_2, \cdots\}$ 是独立同分布随机变量, 其定义如下

$$X_i = \begin{cases} +1, & \text{如果第}i\text{步向右移} \\ -1, & \text{如果第}i\text{步向左移} \end{cases}$$

$$\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = \frac{1}{2}$$

由 $\mathbb{E}[X_i]=0, \mathrm{Var}(X_i)=\mathbb{E}[X_i^2]=1$ 以及(1)式, 可得

$$\begin{aligned}&\mathbb{E}[X(t)]=0\\&\mathrm{Var}[X(t)]=(\Delta x)^2[t/\Delta t]\end{aligned} \tag{2}$$

现在选择 Δx 和 Δt 趋于零的方式。对不同的方式, 极限是不同的。例如, 如果取 $\Delta x=\Delta t$, 那么当 $\Delta t\to 0$ 时, 对任意的 t, $\mathbb{E}[X(t)]=0, \mathrm{Var}[X(t)]\to 0$ 从而 $X(t)\to 0$。

下面取 $\Delta x=\sigma\sqrt{\Delta t}$, $\sigma>0$ 为某常数。从(2)式可知当 $\Delta t\to 0$ 时

$$\mathbb{E}[X(t)]=0$$

$$\mathrm{Var}[X(t)]\to\sigma^2 t$$

以上述方式得到的极限过程仍记为 X, 现在给出 X 的一些直观性质。首先, $X(0)=0$, 另外, 由(1)式和中心极限定理可得

(i) $X(t)$ 服从均值为 0, 方差为 $\sigma^2 t$ 的正态分布。

其次, 由于随机游动在不相交的时间区间中的变化是独立的, 所以

(ii) $\{X(t), t\geqslant 0\}$ 有独立增量性, 即对于所有的 $t_1<t_2<\cdots<t_n, X(t_n)-X(t_{n-1}), X(t_{n-1})-X(t_{n-2}), \cdots, X(t_2)-X(t_1), X(t_1)$ 是相互独立的。

最后, 因为随机游动在任意时间区间中的位置变化的分布只依赖于区间的长度, 从而

(iii) $\{X(t), t\geqslant 0\}$ 有平稳增量性, 即 $X(t+s)-X(t)$ 的分布与 t 无关。

上面给出了一些直观性质, 下面给出 Brown 运动的定义。

定义 如果随机过程 $X=\{X(t), t\geqslant 0\}$ 满足以下性质, 那么就称 X 为 Brown 运动。

(i) $X(0)=0, X$ 的轨道是连续的。

(ii) $\{X(t), t\geqslant 0\}$ 有独立平稳增量性。

(iii) 对每个 $t>0, X(t)\sim N(0,\sigma^2 t)$。

Brown 运动(有时也称为Wiener过程)是应用概率论中最常见的随机过程之一。Brown 运动已经被广泛应用于数学的其他领域, 如微分方程、调和分析等, 它在数理金融中也有重要应用。

特别地, 当 $\sigma=1$ 时, X 称为标准 Brown 运动。对于一般的 Brown 运动 X, 如记 $B(t)=X(t)/\sigma$, 那么 B 就是标准 Brown 运动。因此, 除非特别指明, 通常都假设 $\sigma=1$。



2. 关于 Brown 运动的一些分布函数

设 $X(t) \sim N(0,t)$, $X(t)$ 的密度函数为

$$f_t(x) = \frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-x^2/2t}$$

下面求 $\{X(t_1), X(t_2), \cdots, X(t_n)\}$ 的联合密度函数, 其中 $t_1 < t_2 < \cdots < t_n$。

首先, 注意到以下两组等式是等价的。

$$\{X(t_1) = x_1, X(t_2) = x_2, \cdots, X(t_n) = x_n\}$$

$$\{X(t_1) = x_1, X(t_2) - X(t_1) = x_2 - x_1, \cdots, X(t_n) - X(t_{n-1}) = x_n - x_{n-1}\}$$

其次, 由独立增量性,$X(t_1), X(t_2) - X(t_1), \cdots, X(t_n) - X(t_{n-1})$ 是相互独立的, 另外 $X(t_k) - X(t_{k-1}) \sim N(0, t_k - t_{k-1})$。因此, $X(t_1), X(t_2), \cdots, X(t_n)$ 的联合密度函数是

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) &= f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1) \cdots f_{t_n - t_{n-1}}(x_n - x_{n-1}) \\ &= \frac{\exp\left\{-\dfrac{1}{2}\left[\dfrac{x_1^2}{t_1} + \dfrac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \dfrac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right]\right\}}{(2\pi)^{n/2}[t_1(t_2 - t_1)\cdots(t_n - t_{n-1})]^{1/2}} \end{aligned} \tag{3}$$

由上面的结论, 可以得到另外一些感兴趣的分布。例如, 设 $s < t$, 给定 $X(t) = b$, 可求出 $X(s)$ 的条件分布。下面给出条件密度函数,

$$\begin{aligned} f_{s|t}(x|b) &= \frac{f_s(x) f_{t-s}(b - x)}{f_t(b)} \\ &= K_1 \exp\left\{-\frac{x^2}{2s} - \frac{(b - x)^2}{2(t - s)}\right\} \\ &= K_2 \exp\left\{\frac{1}{2s(t - s)}\left(x^2 - 2\frac{sb}{t}x\right)\right\} \\ &= K_3 \exp\left\{-\frac{(x - bs/t)^2}{2s(t - s)/t}\right\} \end{aligned}$$

其中 K_1, K_2, K_3 是不依赖于 x 的系数。从上面表示可知, $X(s)$ 的条件分布是正态分布, 而且

$$\begin{aligned} \mathbb{E}[X(s)|X(t) = b] &= \frac{s}{t}b \\ \mathrm{Var}[X(s)|X(t) = b] &= \frac{s}{t}(t - s) \end{aligned} \tag{4}$$

例 在甲乙两名选手的自行车比赛中, 以 $Y(t)$ 表示当完成 $100t\%$ 的比赛路程时, 甲领先的时间(以秒计)。假设 $\{Y(t), 0 \leqslant t \leqslant 1\}$ 是方差为 σ^2 的 Brown 运动。

(i) 如果在比赛路程中点时甲领先 σ 秒, 那么甲将赢的概率是多大?

(ii) 如果甲以 σ 秒领先赢得比赛, 那么甲在路程中点时领先的概率是多大?

直接由定义, 可求出第一个概率。

(i)

$$\begin{aligned}\mathbb{P}\{Y(1)>0|Y(1/2)=\sigma\} &= \mathbb{P}\{Y(1)-Y(1/2)>-\sigma|Y(1/2)=\sigma\}\\ &= \mathbb{P}\{Y(1)-Y(1/2)>-\sigma\}=\mathbb{P}\{Y(1/2)>-\sigma\}\\ &= \mathbb{P}\left\{\frac{Y(1/2)}{\sqrt{1/2}\sigma}>-\sqrt{2}\right\}=\Phi(\sqrt{2})\approx 0.9213\end{aligned}$$

其中 $\Phi(x)=\mathbb{P}\{N(0,1)\leqslant x\}$ 是标准正态分布函数。

(ii)

$$\mathbb{P}\{Y(1/2)>0|Y(1)=\sigma\}$$

当 $X(t)=Y(t)/\sigma$ 时, $\{X(t),t\geqslant 0\}$ 是标准 Brown 运动, (4)式给出了 $X(s)$ 的条件分布。因此, 给定 $Y(t)=\sigma X(t)=c$ 时, $Y(s)\sim N(cs/t,\sigma^2 s(t-s)/t)$。最后

$$\begin{aligned}\mathbb{P}\{Y(1/2)>0|Y(1)=\sigma\} &= \mathbb{P}\{N(\sigma/2,\sigma^2/4)>0\}\\ &= \Phi(1)\approx 0.8413\end{aligned}$$

3. 首中时刻

以 T_a 表示 Brown 运动 X 首次到达 a 的时刻。当 $a>0$ 时, 下面计算 $\mathbb{P}\{T_a\leqslant t\}$, 为此考虑 $\mathbb{P}\{X(t)\geqslant a\}$。由全概率公式,

$$\begin{aligned}\mathbb{P}\{X(t)\geqslant a\} =&\mathbb{P}\{X(t)\geqslant a|T_a\leqslant t\}\mathbb{P}\{T_a\leqslant t\}\\ &+\mathbb{P}\{X(t)\geqslant a|T_a>t\}\mathbb{P}\{T_a>t\}\end{aligned}\tag{5}$$

如果 $T_a\leqslant t$, 那么 Brown 运动在时刻 t 以前已经到达 a。由对称性, 在时刻 t, X 等可能地大于 a 或小于 a, 即 $\mathbb{P}\{X(t)\geqslant a|T_a\leqslant t\}=\dfrac{1}{2}$。

另外, 由连续性可知, 当 $t<T_a$ 时, $X(t)<a$ 所以(5)式右边的第二项为零。因此

$$\begin{aligned}\mathbb{P}\{T_a\leqslant t\}=2\mathbb{P}\{X(t)\geqslant a\} &= \frac{2}{\sqrt{2\pi t}}\int_a^\infty \mathrm{e}^{-x^2/2t}\mathrm{d}x\\ &= \frac{2}{\sqrt{2\pi}}\int_{a/\sqrt{t}}^\infty \mathrm{e}^{-y^2/2}\mathrm{d}y\end{aligned}\tag{6}$$

在(6)式中令 $t\to\infty$, 可得 $\mathbb{P}\{T_a<\infty\}=1$。另外, 对(6)式取导数, 即得 T_a 的密度函数的表达式

$$f_{T_a}(t)=\frac{a}{\sqrt{2\pi}}t^{-3/2}\mathrm{e}^{-a^2/2t}$$

从而当 $t \to \infty$ 时，$f_{T_a}(t) \approx \dfrac{a}{\sqrt{2\pi}} t^{-3/2}$，$\mathbb{E}[T_a] = \infty$。

另外，T_1 的 Laplace 变换为

$$\int_0^\infty \frac{1}{\sqrt{2\pi}} t^{-3/2} \mathrm{e}^{-1/2t} \mathrm{e}^{-\alpha t} \mathrm{d}t = \mathrm{e}^{-\sqrt{2\alpha}}, \quad \alpha > 0$$

而 T_a 的 Laplace 变换为 $\mathrm{e}^{-a\sqrt{2\alpha}}$。

当 $a < 0$ 时，由对称性知 T_a 与 $T_{|a|}$ 有相同的分布。从而由(6)式可得

$$\mathbb{P}\{T_a \leqslant t\} = \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^\infty \mathrm{e}^{-y^2/2} \mathrm{d}y \tag{7}$$

如果把 $\{T_a, a \geqslant 0\}$ 视为空间参数为 a 的随机过程，那么 $\{T_a, a \geqslant 0\}$ 具有平稳独立增量性，而且轨道是单增的。

4. 最大值变量

Brown 运动在 $[0,t]$ 内的最大值变量为 $\max\limits_{0 \leqslant s \leqslant t} X(s)$，由(6)式可以很容易求出它的分布。当 $a > 0$ 时，由轨道的连续性，

$$\mathbb{P}\left\{\max_{0 \leqslant s \leqslant t} X(s) \geqslant a\right\} = \mathbb{P}\{T_a \leqslant t\} = \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty \mathrm{e}^{-y^2/2} \mathrm{d}y$$

注意到当 X 是标准 Brown 运动时，$-X$ 也是标准 Brown 运动，而且

$$\max_{0 \leqslant s \leqslant t} \{-X(s)\} = -\min_{0 \leqslant s \leqslant t} \{X(s)\}$$

从而当 $a < 0$ 时，

$$\mathbb{P}\left\{\min_{0 \leqslant s \leqslant t} X(s) \leqslant a\right\} = \mathbb{P}\{T_{|a|} \leqslant t\} = \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^\infty \mathrm{e}^{-y^2/2} \mathrm{d}y$$

5. Brown 运动的零点与 Arcsine 律

设 X 为标准 Brown 运动，对 $x \neq 0$，定义 $X^x(t) = x + X(t)$ 为出发点是 x 的 Brown 运动。

定理 1　X^x 在时间区间 $(0,t)$ 内至少存在一个零点的概率，记为 p，那么

$$p = \frac{|x|}{\sqrt{2\pi}} \int_0^t s^{-3/2} \mathrm{e}^{-x^2/2s} \mathrm{d}s \tag{8}$$

证明　设 $x < 0$，那么

$$\begin{aligned} p &= \mathbb{P}\{\max_{0 \leqslant s \leqslant t} X(s) + x \geqslant 0\} = \mathbb{P}\{\max_{0 \leqslant s \leqslant t} X(s) \geqslant -x\} \\ &= \mathbb{P}\{T_{-x} \leqslant t\} = \mathbb{P}\{T_{|x|} \leqslant t\} = \int_0^t f_{T_{|x|}}(s) \mathrm{d}s = \frac{|x|}{\sqrt{2\pi}} \int_0^t s^{-3/2} \mathrm{e}^{-x^2/2s} \mathrm{d}s \end{aligned}$$

由定理 1 可以得到下面的定理 2。

定理 2 设 $0<a<b$, X 在时间区间 (a,b) 内至少存在一个零点的概率,记为 q,那么

$$q=\frac{2}{\pi}\arccos\sqrt{\frac{a}{b}} \tag{9}$$

证明 引入条件概率 $h(x)$, 它表示在 $X(a)=x$ 的条件下 X 在时间区间 (a,b) 内至少存在一个零点的概率。由 Markov 性, $h(x)$ 与 X^x 在时间区间 $(0,b-a)$ 内至少存在一个零点的概率相同, 即 $h(x)=\frac{|x|}{\sqrt{2\pi}}\int_0^{b-a}s^{-3/2}\mathrm{e}^{-x^2/2s}\mathrm{d}s$。从而

$$q=\int_{-\infty}^{\infty}h(x)\mathbb{P}\{X_a\in\mathrm{d}x\}=\sqrt{\frac{2}{\pi a}}\int_0^{\infty}h(x)\mathrm{e}^{-x^2/2a}\mathrm{d}x$$

把 $h(x)$ 代入上述表达式中, 整理变形就得到定理 2 的结论。详细过程如下

$$\begin{aligned}
q&=\sqrt{\frac{2}{\pi a}}\int_0^{\infty}h(x)\mathrm{e}^{-x^2/2a}\mathrm{d}x=\sqrt{\frac{2}{\pi a}}\int_0^{\infty}\left(\frac{x}{\sqrt{2\pi}}\int_0^{b-a}s^{-3/2}\mathrm{e}^{-x^2/2s}\mathrm{d}s\right)\mathrm{e}^{-x^2/2a}\mathrm{d}x\\
&=\sqrt{\frac{2}{\pi a}}\frac{1}{\sqrt{2\pi}}\int_0^{b-a}s^{-3/2}\mathrm{d}s\left(\int_0^{\infty}x\mathrm{e}^{-x^2/2s}\mathrm{e}^{-x^2/2a}\mathrm{d}x\right)\\
&=\sqrt{\frac{2}{\pi a}}\frac{1}{\sqrt{2\pi}}\int_0^{b-a}s^{-3/2}\left(\frac{1}{s}+\frac{1}{a}\right)^{-1}\mathrm{d}s\\
&=\sqrt{\frac{2}{\pi a}}\frac{1}{\sqrt{2\pi}}(-2)\int_0^{b-a}\left(\frac{1}{s}+\frac{1}{a}\right)^{-1}\mathrm{d}(s^{-1/2})\\
&=\sqrt{\frac{2}{\pi a}}\frac{1}{\sqrt{2\pi}}2\sqrt{a}\arctan(\sqrt{a}s)\Big|_{\frac{1}{\sqrt{b-a}}}^{\infty}\\
&=\sqrt{\frac{2}{\pi a}}\frac{1}{\sqrt{2\pi}}2\sqrt{a}\left(\frac{\pi}{2}-\arctan\frac{\sqrt{a}}{\sqrt{b-a}}\right)\\
&=\frac{2}{\pi}\left(\frac{\pi}{2}-\arctan\frac{\sqrt{a}}{\sqrt{b-a}}\right)=\frac{2}{\pi}\arccos\sqrt{\frac{a}{b}}
\end{aligned}$$

由定理 2 即得下面的定理 3。

定理 3 X 在时间区间 (a,b) 内没有零点的概率为

$$1-q=\frac{2}{\pi}\arcsin\sqrt{\frac{a}{b}} \tag{10}$$

最后, 对给定的 $t>0$, 引入

$$\gamma_t=\sup\{s\leqslant t:X(s)=0\}$$

$$\beta_t=\inf\{s\geqslant t:X(s)=0\}$$

γ_t 表示时刻 t 之前的最后一个零点, β_t 表示时刻 t 之后的第一个零点。那么由前面的定理, 可得到以下定理。

定理 4

$$\mathbb{P}\{\gamma_t \leqslant x\} = \frac{2}{\pi}\arcsin\sqrt{\frac{x}{t}},\ x < t$$

$$\mathbb{P}\{\beta_t \geqslant y\} = \frac{2}{\pi}\arcsin\sqrt{\frac{t}{y}},\ y > t$$

$$\mathbb{P}\{\gamma_t \leqslant x, \beta_t \geqslant y\} = \frac{2}{\pi}\arcsin\sqrt{\frac{x}{y}},\ x < t < y$$

证明 只需注意到 $\{\gamma_t \leqslant x\}$ 等价于 X 在 $[x,t]$ 内没有零点, $\{\beta_t \geqslant y\}$ 等价于 X 在 $[t,y]$ 内没有零点, $\{\gamma_t \leqslant x, \beta_t \geqslant y\}$ 等价于 X 在 $[x,y]$ 内没有零点。

§2 与 Brown 运动有关的过程

1. 有飘移的 Brown 运动

设 $\{B(t), t \geqslant 0\}$ 是标准 Brown 运动, 定义 $X(t) = \sigma B(t) + \mu t$, 称 $\{X(t), t \geqslant 0\}$ 是漂移系数为 μ、方差系数为 σ^2 的 Brown 运动, 它有以下性质

(i) $X(0) = 0, X$ 的轨道是连续的。

(ii) $\{X(t), t \geqslant 0\}$ 有平稳独立增量性。

(iii) $X(t) \sim N(\mu t, \sigma^2 t)$。

2. 几何 Brown 运动

设 $\{Y(t), t \geqslant 0\}$ 是漂移系数为 μ、方差系数为 σ^2 的 Brown 运动, 定义 $X(t) = \mathrm{e}^{Y(t)}$, 称 $\{X(t), t \geqslant 0\}$ 为几何 Brown 运动。

对于几何 Brown 运动 X, 当 $s < t$, 条件期望 $\mathbb{E}[X(t)|X(u), 0 \leqslant u \leqslant s]$ 计算如下

$$\begin{aligned}
\mathbb{E}[X(t)|X(u), 0 \leqslant u \leqslant s] &= \mathbb{E}[\mathrm{e}^{Y(t)}|Y(u), 0 \leqslant u \leqslant s] \\
&= \mathbb{E}[\mathrm{e}^{Y(s)+Y(t)-Y(s)}|Y(u), 0 \leqslant u \leqslant s] \\
&= \mathrm{e}^{Y(s)}\mathbb{E}[\mathrm{e}^{Y(t)-Y(s)}|Y(u), 0 \leqslant u \leqslant s] \\
&= X(s)\mathbb{E}[\mathrm{e}^{Y(t)-Y(s)}]
\end{aligned}$$

设 $W \sim N(\mu, \sigma^2)$, 那么 W 的矩生成函数为

$$\mathbb{E}[\mathrm{e}^{tW}] = \exp\{\mu t + (\sigma^2 t^2/2)\}$$

由于 $Y(t) - Y(s) \sim N(\mu(t-s), (t-s)\sigma^2)$, 应用上述结论即得 $\mathbb{E}[\mathrm{e}^{Y(t)-Y(s)}] = \mathrm{e}^{\mu(t-s)+(t-s)\sigma^2/2}$。最后得到

$$\mathbb{E}[X(t)|X(u), 0 \leqslant u \leqslant s] = X(s)\mathrm{e}^{(t-s)(\mu+\sigma^2/2)}$$

特别地, 如果 $\mu + \sigma^2/2 = 0$ 那么

$$\mathbb{E}[X(t)|X(u), 0 \leqslant u \leqslant s] = X(s)$$

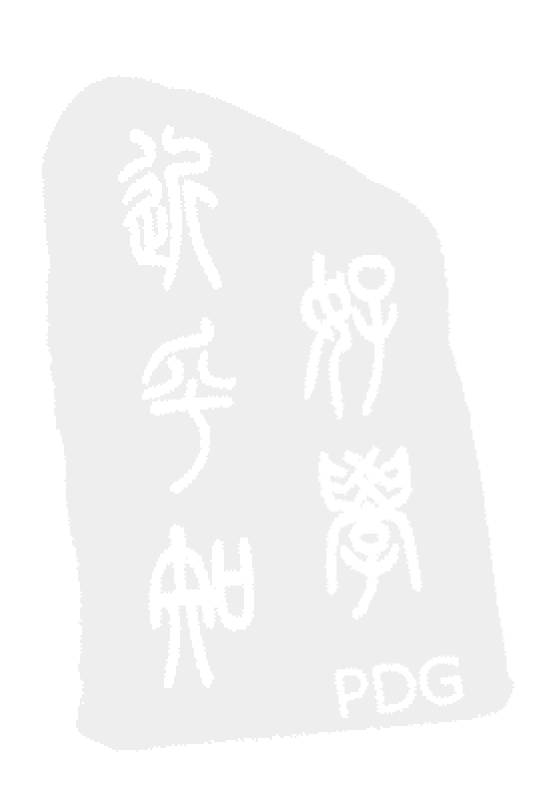
如果经验表明股票价格的变化率是独立同分布的, 那么几何 Brown 运动就是很有用的股票价格模型。例如, 假设 X_n 是时刻 n 时某股票的价格, 而且 $\{X_n/X_{n-1}, n \geqslant 1\}$ 是独立同分布的。记 $Y_n = X_n/X_{n-1}$, 那么 $X_n = Y_n X_{n-1}$。依次类推, 可得

$$X_n = Y_n Y_{n-1} X_{n-2} = Y_n Y_{n-1} Y_{n-2} X_{n-3} = \cdots = Y_n Y_{n-1} \cdots Y_1 X_0$$

因此,

$$\log(X_n) = \sum_{i=1}^{n} \log(Y_i) + \log(X_0)$$

因为 $\{\log(Y_i), i \geqslant 1\}$ 是独立同分布的, 所以当 n 较大时, 上式右边第一项近似于某个正态分布变量 $Y_n, Y_n \sim N(n\mu, n\sigma^2)$, 从而 $X_n = X_0 \mathrm{e}^{Y_n}$, 这就解释了价格过程 X 可近似地用几何 Brown 运动来描述。



第四章　随机过程的公理化定义

§1　概率空间

1. 集合论中的一些基本概念

设 Ω 是某集合或空间, 其中的一般元素记为 ω。如果集合 A 的所有元素都属于 Ω, 那么称 A 为 Ω 的子集, 记为 $A \subset \Omega$。

首先给出一些标准的集合运算及其规则。集合 A 的余集表示一个集合, 该集合包含了 Ω 中所有不属于 A 的元素, 即

$$A^c = \{\omega; \omega \notin A\}$$

对一族集合 $\{A_\gamma; \gamma \in \Gamma\}$, 它们的并表示一个集合, 该集合包含了所有至少属于一个集合 A_γ 的元素, 即

$$\bigcup_{\gamma \in \Gamma} A_\gamma = \{\omega \in \Omega; 存在某个\gamma \in \Gamma, 使得\omega \in A_\gamma\}$$

另外, 对一族集合 $\{A_\gamma; \gamma \in \Gamma\}$, 它们的交表示一个集合, 该集合包含了属于所有集合 A_γ 的元素, 即

$$\bigcap_{\gamma \in \Gamma} A_\gamma = \{\omega \in \Omega; 对每个\gamma \in \Gamma, \omega \in A_\gamma\}$$

关于集合的并与交, 下面的 De Morgan 律成立,

$$\Big(\bigcap_{\gamma \in \Gamma} A_\gamma\Big)^c = \bigcup_{\gamma \in \Gamma} A_\gamma^c \tag{1}$$

(1)式的解释就是如果一个元素不属于所有的集合 A_γ, 那么该元素一定在某个集合 A_γ 外。

引入空集 $\varnothing$, 它不包含任何元素。由定义, 空集可视为所有其他集合的子集, 而且 $\varnothing = \Omega^c$。如果 $A \bigcap B = \varnothing$, 那么称集合 A 和 B 是不相交的。集合的差运算定义为

$$A \backslash B = A \cap B^c = \{\omega; \omega \in A, \omega \notin B\}$$

2. 概率空间的定义

概率用来表示不确定性程度。在不确定性存在的情形下, 概率论是描述与分析问题的数学方法。

首先, 确定样本空间 Ω, 它是所有可能出现的结果 ω 的集合。其次, 确定 Ω 中的某些集合, 它们表示某些感兴趣的事件。这些集合或事件的集合 $\mathcal{F}$ 应包含 Ω(必然事件), 而且, $\mathcal{F}$ 关于余集运算与可列个集合的并运算应有封闭性。

定义 1 如果集合 $\mathcal{F}$ 具有以下性质

(i) $\Omega \in \mathcal{F}$。

(ii) 如果 $A \in \mathcal{F}$, 那么 $A^c \in \mathcal{F}$。

(iii) 对 $\mathcal{F}$ 内的可列集合 $\{A_j\}_{j=1,2,\cdots}$, $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$。

那么称集合 $\mathcal{F}$ 是 Ω 上的 σ– 代数。

称 $(\Omega, \mathcal{F})$ 为可测空间, 事件 $A \in \mathcal{F}$ 也称为可测集。对每个事件 $A \in \mathcal{F}$, 赋予 [0,1] 区间上的一个数值, 称为 A 的概率, 记为 $\mathbb{P}[A]$。必然事件 Ω 的概率应为 1, 而且互不相交的事件的并的概率应为各个事件的概率之和。

定义 2 设 $(\Omega, \mathcal{F})$ 为可测空间。如果集函数 $\mathbb{P}: \mathcal{F} \to [0,1]$ 具有以下性质

(i) $\mathbb{P}[\Omega] = 1$。

(ii) 对 $\mathcal{F}$ 内互不相交的可列集合 $\{A_j\}_{j=1,2,\cdots}$, $\mathbb{P}\left[\bigcup_{j=1}^{\infty} A_j\right] = \sum_{j=1}^{\infty} \mathbb{P}[A_j]$。

那么称 $\mathbb{P}$ 为 $(\Omega, \mathcal{F})$ 上的概率测度。

称 $(\Omega, \mathcal{F}, \mathbb{P})$ 为概率空间。如果对 $A \in \mathcal{F}, \mathbb{P}[A] = 0$, 那么称 A 为零集合。如果某性质对 Ω 内某个零集合以外的所有 ω 成立, 那么称其为几乎处处成立。

3. 概率空间的一般性质

由定义 1 的 (i) 和 (ii) 可知 $\varnothing \in \mathcal{F}$。由定义 2 的 (i) 和 (ii) 可知 $\mathbb{P}[\varnothing] = 0$。由 (1) 式和定义 1 的 (ii) 和 (iii) 可知 $\mathcal{F}$ 关于可列事件的交运算也是封闭的。对任意

事件 A 和 B, 下式成立

$$\mathbb{P}[A\cup B]=\mathbb{P}[A]+\mathbb{P}[B]-\mathbb{P}[A\cap B]$$

如果 $A\subset B$, 那么 $\mathbb{P}[A]\leqslant\mathbb{P}[B]$, 而且 $\mathbb{P}[B\backslash A]=\mathbb{P}[B]-\mathbb{P}[A]$。

对一列事件 $\{A_j\}_{j=1,2,\cdots}$, 下面的次可加性是成立的

$$\mathbb{P}\Big[\bigcup_{j=1}^{\infty}A_j\Big]\leqslant\sum_{j=1}^{\infty}\mathbb{P}[A_j] \tag{2}$$

把集合的并分解为一列互不相交的集合的并得到, 即

$$\bigcup_{j=1}^{\infty}A_j=\bigcup_{j=1}^{\infty}(A_j\cap A_0^c\cap\cdots\cap A_{j-1}^c)$$

在上式中定义 $A_0=\varnothing$, 然后应用定义 2 的 (ii) 可得到 (2) 式。

引理 1 (概率的连续性)　如果 $\{A_j\}_{j=1,2,\cdots}$ 是单调的(递增或递减), 那么

$$\mathbb{P}[\lim_{n\to\infty}A_n]=\lim_{n\to\infty}\mathbb{P}[A_n]$$

此时, 如果 $\{A_j\}_{j=1,2,\cdots}$ 是递增的, $\lim\limits_{n\to\infty}A_n=\bigcup\limits_{j=1}^{\infty}A_j$; 如果 $\{A_j\}_{j=1,2,\cdots}$ 是递减的, $\lim\limits_{n\to\infty}A_n=\bigcap\limits_{j=1}^{\infty}A_j$。

特别地, 如果 $\{A_j\}_{j=1,2,\cdots}$ 是递减的, 而且 $\bigcap\limits_{j=1}^{\infty}A_j=\varnothing$, 那么 $\lim\limits_{n\to\infty}\mathbb{P}[A_n]=0$。

证明　首先, 设 $\{A_j\}_{j=1,2,\cdots}$ 是递增的, 引入 $A_0=\varnothing$, 由定义 2 的 (ii),

$$\begin{aligned}\mathbb{P}\left[\bigcup_{j=1}^{\infty}A_j\right]&=\mathbb{P}\left[\bigcup_{j=1}^{\infty}(A_j\backslash A_{j-1})\right]=\sum_{j=1}^{\infty}(\mathbb{P}[A_j]-\mathbb{P}[A_{j-1}])\\&=\lim_{n\to\infty}\sum_{j=1}^{n}(\mathbb{P}[A_j]-\mathbb{P}[A_{j-1}])=\lim_{n\to\infty}\mathbb{P}[A_n]\end{aligned}$$

其次, 当 $\{A_j\}_{j=1,2,\cdots}$ 递减时, 由 (1) 式和上面的结论, 即得引理结论也成立。

对任意的一列集合, 定义下极限为一个集合, 其中的元素属于有限个集合外的所有集合, 记为

$$\liminf A_j=\bigcup_{j=1}^{\infty}\bigcap_{k=j}^{\infty}A_k=\lim_{j\to\infty}\bigcap_{k=j}^{\infty}A_k$$

在上式中用到 $\{\bigcap_{k=j}^{\infty} A_k\}, j=1,2,\cdots$ 是递增的性质。类似地，定义上极限为一个集合，其中的元素属于无限多个集合，记为

$$\limsup A_j = \bigcap_{j=1}^{\infty}\bigcup_{k=j}^{\infty} A_k = \lim_{j\to\infty}\bigcup_{k=j}^{\infty} A_k$$

显然，$\liminf A_j \subset \limsup A_j$。应用于事件时，一列集合的上极限和下极限的直观解释就是：$\liminf A_j$ 表示一个事件，此时，除了有限个事件 A_j 不发生外，所有的事件 A_j 都发生。$\limsup A_j$ 表示另一个事件，此时有无穷多个事件 A_j 发生。

引理 2 (Borel–Cantelli 第一引理) 如果

$$\sum_{j=1}^{\infty}\mathbb{P}[A_j] < \infty \tag{3}$$

那么 $\mathbb{P}[\limsup A_j] = 0$。

证明

$$\mathbb{P}[\limsup A_j] = \mathbb{P}\left[\lim_{j\to\infty}\bigcup_{k=j}^{\infty} A_k\right] = \lim_{j\to\infty}\mathbb{P}\left[\bigcup_{k=j}^{\infty} A_k\right] \leqslant \lim_{j\to\infty}\sum_{k=j}^{\infty}\mathbb{P}[A_k] = 0$$

在上式中用到次可加性 (2) 式以及条件 (3) 式。

§2 随机变量与条件期望

1. 随机变量与期望

给定概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 以及函数 $X:\Omega\to\mathbb{R}$，如果对 $\mathbb{R}$ 中任意 Borel 集 B，$X^{-1}(B)=\{\omega; X(\omega)\in B\}\in\mathcal{F}$，那么称 X 是可测的，或称 X 为随机变量 (r.v.)，有时简记为 $X\in\mathcal{F}$。X 的累积分布函数 (c.d.f.) 记为 F_X，$F_X(x)=\mathbb{P}\{\omega; X(\omega)\leqslant x\}$。

随机变量 X 的期望的定义为

$$\mathbb{E}[X] = \int X\mathrm{d}\mathbb{P} = \int_{\Omega} X(\omega)\mathrm{d}\mathbb{P}(\omega)$$

对给定的随机变量 X，由所有的集合 $X^{-1}(B)=\{\omega; X(\omega)\in B\}$ 可以构成 $\mathcal{F}$ 的一个子 σ- 代数，记为 $\mathcal{G}$，那么关于概率空间 $(\Omega,\mathcal{G},\mathbb{P})$，$X$ 也是随机变量，而且 $\mathcal{G}$ 是使 X 成为随机变量的最小 σ- 代数。有时候，也记 $\mathcal{G}=\sigma(X)$。

2. 条件期望

设 $\mathcal{G}$ 是 $\mathcal{F}$ 的子 σ- 代数, 记限于 $\mathcal{G}$ 上的 $\mathbb{P}$ 为 $\mathbb{P}_{\mathcal{G}}$, 从而对 $A\in\mathcal{G}$, $\mathbb{P}_{\mathcal{G}}[A]=\mathbb{P}[A]$ 成立。显然 $(\Omega,\mathcal{G},\mathbb{P}_{\mathcal{G}})$ 是概率空间。对随机变量 Y, 如果对任意 Borel 集 B,$Y^{-1}(B)\in\mathcal{G}$, 那么称 Y 是关于 $\mathcal{G}$ 可测的, 简记为 $Y\in\mathcal{G}$。

设随机变量 X 的期望存在, X 关于 $\mathcal{G}$ 的条件期望是一个随机变量, 记为 $\mathbb{E}[X|\mathcal{G}]$, $\mathbb{E}[X|\mathcal{G}]\in\mathcal{G}$, 而且对所有的 $Y\in\mathcal{G}$, 下式成立

$$\mathbb{E}\{\mathbb{E}[X|\mathcal{G}]Y\}=\mathbb{E}[XY] \tag{4}$$

特别地, 如在 (4) 式中取 $Y\equiv 1$, 可得到如下关于条件期望的迭代性质

$$\mathbb{E}\{\mathbb{E}[X|\mathcal{G}]\}=\mathbb{E}[X]$$

设 X 和 Y 是两个随机变量, 而且 $\mathcal{G}=\sigma(Y)$, 有时候, $\mathbb{E}[X|\mathcal{G}]$ 也写为 $\mathbb{E}[X|Y]$。

条件期望还有一个重要的性质, 可称为塔式性质, 即下面的引理。

引理 3 设 $\mathcal{H}\subset\mathcal{G}\subset\mathcal{F}$, 那么

$$\mathbb{E}\{\mathbb{E}[X|\mathcal{G}]|\mathcal{H}\}=\mathbb{E}[X|\mathcal{H}] \tag{5}$$

例 1 为理解 (4) 式, 考虑一个特殊情形, $\mathcal{G}=\sigma\{B_1,B_2,\cdots\}$, 其中 $\mathcal{F}$ 可测集 $B_1,B_2,\cdots$ 构成了 Ω 的一个划分, 即 $\{B_j\}_{j=1,2,\cdots}$ 互不相交, 而且 $\bigcup\limits_{j=1}^{\infty}B_j=\Omega$。$\mathbb{E}[X|\mathcal{G}]$ 是 $\mathcal{G}$ 可测的, 所以 $\mathbb{E}[X|\mathcal{G}]$ 的形式为 $\sum\limits_k b_k 1_{B_k}$。在 (4) 式中取 $Y=1_{B_j}$, 得到

$$\mathbb{E}[X|\mathcal{G}]=\sum_j 1_{B_j}\frac{\int_{B_j}X\mathrm{d}\mathbb{P}}{\mathbb{P}[B_j]}$$

特别地, 取 $X=1_A$, 可得

$$\mathbb{E}[1_A|\mathcal{G}]=\sum_j 1_{B_j}\frac{\mathbb{P}[A\cap B_j]}{\mathbb{P}[B_j]}$$

上式与常见的条件概率的定义 $\mathbb{P}[A|B]=\mathbb{P}[A\cap B]/\mathbb{P}[B]$ 相一致, 差别在于这里的条件期望是随机变量。有时候, $\mathbb{E}[1_A|\mathcal{G}]$ 又记为 $\mathbb{P}[A|\mathcal{G}]$, 称为条件概率。

例 2 此例引入条件密度的概念。设 $f(\omega_1,\omega_2)$ 是定义在 $\Omega=\mathbb{R}^2$ 上的密度函数, 考虑随机变量 $X(\omega_1,\omega_2)=\omega_1$ 和 $Y(\omega_1,\omega_2)=\omega_2$, 它们都是坐标函数。记 $\mathcal{G}=\sigma(Y)=\{\mathbb{R}\times B;B\in\mathcal{B}\}$,$\mathcal{B}$ 是 $\mathbb{R}$ 上的 Borel 集构成的 σ- 代数。每个 $\mathcal{G}$ 可测函数的形式为 $g(\omega_2)$。特别地, 记 $\mathbb{E}[X|\mathcal{G}]=h(\omega_2)$。(4) 式表明对所有的 g, 下式成立

$$\iint h(\omega_2)g(\omega_2)f(\omega_1,\omega_2)\mathrm{d}\omega_1\mathrm{d}\omega_2 = \iint \omega_2 g(\omega_2)f(\omega_1,\omega_2)\mathrm{d}\omega_1\mathrm{d}\omega_2$$

对上式变形, 得到

$$\int g(\omega_2)h(\omega_2)\left(\int f(\omega_1,\omega_2)\mathrm{d}\omega_1\right)\mathrm{d}\omega_2 = \int g(\omega_2)\left(\int \omega_1 f(\omega_1,\omega_2)\mathrm{d}\omega_1\right)\mathrm{d}\omega_2$$

从而得到

$$h(\omega_2) = \int \omega_1 f(\omega_1|\omega_2)\mathrm{d}\omega_1$$

其中 $f(\omega_1|\omega_2)$ 是通常的在给定 Y 下 X 的条件密度,

$$f(\omega_1|\omega_2) = \frac{f(\omega_1,\omega_2)}{f_Y(\omega_2)}, f_Y(\omega_2) = \int f(\omega_1,\omega_2)\mathrm{d}\omega_1$$

3. 独立性

如果下式成立

$$\mathbb{P}[A\cap B] = \mathbb{P}[A]\mathbb{P}[B] \tag{6}$$

那么称事件 A 和 B 关于 $\mathbb{P}$ 是独立的。进一步, 对 $\mathcal{F}$ 的两个子 σ- 代数 $\mathcal{G}$ 和 $\mathcal{H}$, 如果对任意的 $A\in\mathcal{G}$ 和 $B\in\mathcal{H}$, (6)式成立, 那么称 $\mathcal{G}$ 和 $\mathcal{H}$ 关于 $\mathbb{P}$ 是独立的。对 $X\in\mathcal{G}$ 和 $Y\in\mathcal{H}$, 设 $\mathcal{G}$ 和 $\mathcal{H}$ 关于 $\mathbb{P}$ 是独立的, 那么 (6) 式有如下推广

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

把上式代入 (4) 式中, 并取 Y 为常数, 可得

$$\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$$

§3 构造特殊的概率空间

1. 确定事件与概率

确定与 §1 中定义 1 相一致的适当的事件集合 $\mathcal{F}$, 以及与 §1 中定义 2 相一致的适当的集函数 $\mathbb{P}$, 并不总是简单的问题。这依赖于样本空间 Ω 的性质, 以及要求概率空间所具有的性质。

如果样本空间 Ω 的元素是有限的或可列的, $\Omega=\{\omega_1,\omega_2,\cdots\}$, 那么选取一列非负数 $\{p_1,p_2,\cdots\}$, 使得 $\sum\limits_{j=1}^{\infty}p_j=1$, 而对每个 $A\subset\Omega$, 定义

$$\mathbb{P}[A] = \sum_{j;\omega_j\in A} p_j$$

那么容易验证 $\mathbb{P}$ 是定义于由 Ω 的所有子集构成的集合 $\mathcal{F}$ 上的概率测度。

如果样本空间 Ω 的元素是不可列的, 那么可测性是一个需要考虑的问题。例如, 当 $\Omega=\mathbb{R}$ 时, 如何构造适当的 σ- 代数并不是显然的。一个很自然的出发点就是累积分布函数。首先确定一个非减的、右连续函数 F, 并要求 $\lim\limits_{x\to-\infty}F(x)=0$, $\lim\limits_{x\to\infty}F(x)=1$。其次定义半开半闭区间上的概率 $\mathbb{P}[(x,y]]=F(y)-F(x)$, 开区间上的概率 $\mathbb{P}[(x,y)]=F(y-)-F(x)$, 点上的概率 $\mathbb{P}[\{x\}]=\mathbb{P}[[x,x]]=F(x)-F(x-)$, 以及可列个区间之并上的概率。也许从实用的角度讲, 可列个区间之并所构成的集合类已经是很广的一类集合, 从而不必考虑这类集合以外的集合的概率。但至此仍有一个问题: 适当的 σ- 代数是什么? 事实上, 可列个区间之并所构成的集合类并不能构成一个 σ- 代数。例如, 有理数是可列个区间(退化的闭区间)之并, 但有理数的余集, 即无理数显然不能表示为可列个区间之并。

2. 存在性定理

可测性问题可按下面的一般方案来解决: 首先考虑一类简单集合类 $\mathcal{C}$(如直线上的区间集合), 并定义简单集合的概率, 它是一个集函数, 记为 $\mathbb{P}_0$。集合类 $\mathcal{C}$ 构成一个代数, 即它关于余集运算以及有限个集合的并运算封闭。另外, 要求定义的集函数 $\mathbb{P}_0$ 满足 §1 中定义 2 的 (i) 和 (ii)。至此, 应用测度论的如下基本结论可以构造概率空间 $(\Omega,\mathcal{F},\mathbb{P})$。

定理 1 设 $\mathcal{F}=\sigma(\ )$ 是包含 $\mathcal{C}$ 的最小的 σ- 代数, $\mathbb{P}_0$ 可唯一扩充成 $\mathcal{F}$ 上的概率测度 $\mathbb{P}$。

3. 有限维欧几里得空间上的概率

前面关于实数空间上的概率测度的讨论可以推广到一般的有限维欧几里得空间 $\mathbb{R}^d$, 此时仍然从区间出发。通常, 累积分布函数 F 由密度函数 $f:\mathbb{R}^d\to\mathbb{R}_+$ 给出,

$$F(x_1,\cdots,x_d)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d}f(\omega_1,\cdots,\omega_d)\mathrm{d}\omega_1\cdots\mathrm{d}\omega_d$$

4. 函数空间上的概率

设 $(X_t)_{t\geqslant0}$ 是实值函数, 它表示某些过程(例如股票价格或利率)的随机变化。此时面临的问题是如何以相容的方式在函数空间 $\mathbb{R}^{\mathbb{R}_+}$ 的某个 σ- 代数上定义概率测度。对有限个时刻 $t_1,\cdots,t_d$, 考虑 $X_{t_1},\cdots,X_{t_d}$ 的累积分布函数 $F_{t_1,\cdots,t_d}$, 要求 $F_{t_1,\cdots,t_d}$ 满足相容性, 即低维的累积分布函数是高维的累积分布函数的边际。按照这种方式, 对 $\mathbb{R}^d$ 中任意 Borel 集 B, 定义在由形如 $\{\omega=(x_t)_{t\geqslant0};(x_{t_1},\cdots,x_{t_d})\in B\}$ 的柱集类构成的代数上的集函数满足 §1 中定义 2 的 (i) 和 (ii)。进一步可推广到包含柱集类的最小 σ- 代数上的概率测度。

如果时间参数是离散的, 如 $t \in \{0,1,2,\cdots\}$, 可按以下条件概率密度方式递推构造有限维分布

$$f_{0,\cdots,d-1,d}(x_0,\cdots,x_{d-1},x_d) = f_{d|0,\cdots,d-1}(x_d|x_0,\cdots,x_{d-1})f_{0,\cdots,d-1}(x_0,\cdots,x_{d-1})$$

上式表明, 为构造有限维分布, 首先要确定 X_0 的边际密度, 以及对每个 d, 给定 $X_0,\cdots,X_{d-1}$ 下 X_d 的条件分布。这样构造的有限维分布自然就满足相容性要求。另外, 由于条件密度是容易理解的概念, 上述构造也是一种非常自然的方式。例如, 如果要求所构造的过程具有 Markov 性, 那么须要求 $f_{d|0,\cdots,d-1}(x_d|x_0,\cdots,x_{d-1}) = f_d(x_d|x_{d-1})$; 而如果要求所构造的过程具有独立性, 那么须要求 $f_{d|0,\cdots,d-1}(x_d|x_0,\cdots,x_{d-1}) = f_d(x_d)$。

上述思路不能直接应用于时间参数为连续的情形, 但仍有所启示。例如, 在跳跃过程中, 通过引入密度函数, 可以确定很短的时间区间上的条件分布。

上述讨论与随机过程的构造有密切的关系。

5. 完备概率空间

从前述讨论中, $\mathcal{F}$ 的定义出现于 $\mathbb{P}$ 之前。但是, 出于一些技术处理的方便需要, 可以假设概率空间是完备的, 即 $\mathcal{F}$ 包含了所有 $\mathbb{P}$ 零集及其子集。因此, 如果开始构造的概率空间 $(\Omega,\mathcal{F}_0,\mathbb{P}_0)$ 不是完备的, 那么可以按照以下方式构造对应的完备空间 $(\Omega,\mathcal{F},\mathbb{P})$: 令 $\mathcal{F}$ 包含所有形如 $A = A_0 \cup N$ 的集合, 其中 $A_0 \in \mathcal{F}_0, N \subset B \in \mathcal{F}_0$, 而且 $\mathbb{P}_0[B] = 0$。对上述集合 A, 定义 $\mathbb{P}[A] = \mathbb{P}_0[A_0]$。容易验证 $\mathcal{F}$ 是一个 σ- 代数, 它包含了 $\mathcal{F}_0$(只须令 $N = \varnothing$ 即知), 而且 $\mathbb{P}$ 和 $\mathbb{P}_0$ 在 $\mathcal{F}_0$ 上是重合的。

§4 随机过程

1. 过滤的概率空间

为了描述在时间区间 $[0,T]$ 上随机现象的演化, 引入 $\mathcal{F}$ 内一族子 σ- 代数, 记为 $\mathbf{F} = \{\mathcal{F}_t\}_{0\leqslant t\leqslant T}$, 其中 $\mathcal{F}_t$ 表示到时刻 t 时已有的信息。准确地讲, $\mathcal{F}_t$ 包含的事件是否发生可在时刻 t 确定。在不失去信息的前提下, 对 $s < t$, 应有 $\mathcal{F}_s \subset \mathcal{F}_t$。称 $\mathbf{F}$ 为滤波, 而 $(\Omega,\mathcal{F},\mathbf{F},\mathbb{P})$ 为过滤的概率空间。有时候, 要求 $\mathbf{F}$ 是右连续的, 即对任意 t,

$$\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$$

而且 $\mathcal{F}_0$(从而 $\mathcal{F}_t$)是完备的。

2. 随机过程

随机过程由一族随机变量构成, 即 $X = \{X_t\}_{0\leqslant t\leqslant T}$。对每个 $\omega \in \Omega$, 称 $\{X_t(\omega)\}_{0\leqslant t\leqslant T}$ 为随机过程的一条路径或样本轨道。如果对每个 $t \in [0,T], X_t \in \mathcal{F}_t$, 那么称随机过程关于滤波 $\mathbf{F}$ 是可适应的, 即在每一时刻, 如果已知 $\mathbf{F}$, 那么过程的当前状态(以及过去的历史)也就完全知道了。进一步, 对一个可适应的过程, 如果在每一时刻的取值可完全由过程的严格过去历史来确定, 那么称该过程为可预料的。严格过去历史记为 $\mathcal{F}_{t-} = \bigcup\limits_{\varepsilon>0} \mathcal{F}_{t-\varepsilon}$。出于以后应用的需要, 这里要指出, 如果过程轨道是左连续的, 或者是确定的, 那么该过程就是可预料的。

3. Markov 链

取值于离散状态的随机过程 X, 如果满足下列关系

$$\mathbb{P}[X_t = x|\mathcal{F}_s] = \mathbb{P}[X_t = x|X_s], s \leqslant t$$

那么称 X 为 Markov 链。上述定义表明, 在给定过去的信息这一条件下, 随机过程的未来取值的条件概率仅依赖于最近的信息。特别地, 如果 $\{X_n\}_{n=1}^{\infty}$ 是 Markov 链, 设 $\mathcal{F}_n = \sigma(X_1, X_2, \cdots, X_n)$, 那么

$$\mathbb{E}[X_{n+m}|\mathcal{F}_n] = \mathbb{E}[X_{n+m}|X_1, X_2, \cdots, X_n] = \mathbb{E}[X_{n+m}|X_n]$$

4. 鞅

对具有有限期望的可适应的过程 X, 如果对 $s < t$, 下式成立

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s$$

那么称 X 为鞅。鞅性质既依赖于滤波, 也依赖于概率测度。特别地, $\mathbb{E}[X_t|X_s] = X_s$, 它表明在平均意义下, 一个鞅的未来取值的平均总是保持在当前水平上。容易验证, 在给定当前信息的条件下, 鞅在将来不相交时间区间上的增量是不相关的, 即对$s \leqslant t \leqslant u \leqslant v, \mathrm{Cov}(X_t - X_s, X_v - X_u|\mathcal{F}_s) = 0$。给定可积随机变量 Y, 构造 $X_t = \mathbb{E}[Y|\mathcal{F}_t]$, 那么由 §2 节的 (5) 式, 可以验证随机过程 $\{X_t\}_{t\geqslant 0}$ 是一个鞅。

5. 停时

停时的概念经常于鞅论研究中。停时是特殊的非负随机变量 T, 使得对任意的 $t \geqslant 0, \{\omega : T(\omega) \leqslant t\} \in \mathcal{F}_t$。停时的直观解释是: 是否在时刻 t 停止, 只须在直到时刻 t 时的信息基础上即可做出这一决策, 即 $\{\omega : T(\omega) \leqslant t\} \in \mathcal{F}_t, \{\omega : T(\omega) > t\} \in \mathcal{F}_t$。

例 3 设 a 属于可适应过程 X 的取值范围, 定义随机变量如下

$$T_1 = \inf\{t : X_t = a\}$$

$$T_2 = \sup\{t : X_t = a\}$$

那么 T_1 是停时, 但 T_2 不是停时。直观地讲,$\{T_1 > t\}$ 等价于在时刻 t 前 $\{X_s < a\}$ 成立, 由可适应性的假设,$\{X_s < a\} \in \mathcal{F}_t$。但对 T_2 来说, 即使在某时刻 t, $X_t = a$, 但为了验证时刻 t 是最后到达时刻, 需要观察时刻 t 以后的信息, 所以一般地 $\{T_2 \leqslant t\} \notin \mathcal{F}_t$。

6. 计数过程

计数过程是一类简单的随机过程 $N = \{N_t\}_{0\leqslant t\leqslant T}$, 它的初始值为 0, 即 $N_0 = 0$, 以后每次跳跃度都是 1。记 N 的自然滤波为 $\mathbf{F}^N = \{\mathcal{F}_t^N\}_{0\leqslant t\leqslant T}$, 其中 $\mathcal{F}_t^N = \sigma\{N_s; s \leqslant t\}$ 是 N 到时刻 t 时的历史, N 关于自然滤波是可适应的, 而且自然滤波是最小的滤波, 使得 N 是可适应的。N 在时刻 t 时的严格过去历史记为 $\mathcal{F}_{t-}^N$。

设 $\{\Lambda_t\}_{0\leqslant t\leqslant T}$ 是 $\mathbf{F}^N$ 可预料过程, 构造如下过程 M

$$M_t = N_t - \Lambda_t \tag{7}$$

如果 M 是一个零均值的 $\mathbf{F}^N$ 鞅, 那么称 $\{\Lambda_t\}_{0\leqslant t\leqslant T}$ 是 N 的补偿。进一步, 如果 Λ 是绝对连续的, 即

$$\Lambda_t = \int_0^t \lambda_s \mathrm{d}s$$

那么称过程 λ 为 N 的强度。形式上强度 λ 有如下表示

$$\lambda_t \mathrm{d}t = \mathbb{P}[\mathrm{d}N_t = 1|\mathcal{F}_{t-}] = \mathbb{E}[\mathrm{d}N_t|\mathcal{F}_{t-}]$$

由 (7) 式给出的鞅的微分形式为

$$\mathrm{d}M_t = \mathrm{d}N_t - \lambda_t \mathrm{d}t \tag{8}$$

设 h 是 $\mathbf{F}^N$ 可预料过程, h 关于鞅 M 的随机积分记为 $\int_0^t h_s \mathrm{d}M_s$, 它是 $\mathbf{F}^N$ 可适应过程。而且当 h 满足一定条件时, 随机积分是一个鞅。

一个基本表示结论表明每一个 $\mathbf{F}^N$ 鞅都可以表示为关于 M 的随机积分。设 H 是一个鞅, 那么存在 $\mathbf{F}^N$ 可预料过程 h, 使得 H 可表示为如下形式,

$$H_t = H_0 + \int_0^t h_s \mathrm{d}M_s \tag{9}$$

其中 $H_0 \in \mathcal{F}_0^N$。

设 $H_t^{(1)} = H_0^{(1)} + \int_0^t h_s^{(1)} \mathrm{d}M_s, H_t^{(2)} = H_0^{(2)} + \int_0^t h_s^{(2)} \mathrm{d}M_s$, 如果 $H_t^{(1)}$ 和 $H_t^{(2)}$ 都有有限方差, 那么

$$\mathrm{Cov}[H_T^{(1)}, H_T^{(2)}|\mathcal{F}_t] = \mathbb{E}\left[\int_t^T h_s^{(1)} h_s^{(2)} \lambda_s \mathrm{d}s|\mathcal{F}_t\right] \tag{10}$$

特别地,

$$\mathrm{Var}[H_T|\mathcal{F}_t]=\mathbb{E}\left[\int_t^T h_s^2\lambda_s\mathrm{d}s|\mathcal{F}_t\right]$$

如果 $H^{(1)}$ 和 $H^{(2)}$ 在时间区间 $[t,s]$ 上的增量为零, 那么称 $H^{(1)}$ 和 $H^{(2)}$ 是正交的, 它等价于 $H^{(1)}\times H^{(2)}$ 是一个鞅。

称强度过程 λ 为计数过程 N 的无穷小特征, λ 完全决定了 N 的概率性质。如果 λ 是确定性的, 那么 N 是 Poisson 过程。如果 λ 只依赖于 N_{t-}, 那么 N 是一个 Markov 过程。

§5 测度变换

1. Radon-Nikodym 定理

设 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$ 是定义于同一个可测空间 $(\Omega,\mathcal{F})$ 上的概率测度, 如果所有 $\mathbb{P}$ 零集都是 $\tilde{\mathbb{P}}$零集, 那么称 $\tilde{\mathbb{P}}$ 关于 $\mathbb{P}$ 是绝对连续的, 记为 $\tilde{\mathbb{P}}\ll\mathbb{P}$。如果 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$ 是相互绝对连续的, 那么称 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$ 是等价的, 记为 $\tilde{\mathbb{P}}\sim\mathbb{P}$, 此时 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$ 有相同的零集。

如果 L 是随机变量, 使得 $L\geqslant 0$, 而且 $\mathbb{E}[L]=1$, 定义 $\mathcal{F}$ 上的概率测度 $\tilde{\mathbb{P}}$ 如下,

$$\tilde{\mathbb{P}}[A]=\int_A L\mathrm{d}\mathbb{P}=\mathbb{E}[1_A L] \tag{11}$$

显然, $\tilde{\mathbb{P}}\ll\mathbb{P}$。Radon-Nikodym 定理表明, 如果 $\tilde{\mathbb{P}}\ll\mathbb{P}$, 那么存在随机变量 L, 使得对所有的 $A\in\mathcal{F}$, (11) 式成立, 此时称 L 为 $\tilde{\mathbb{P}}$ 关于 $\mathbb{P}$ 的导数。(11) 式的微分形式为

$$\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}=L \tag{12}$$

2. 测度变换下的性质

性质 1 设两个测度 $\mathbb{P}$ 和 $\tilde{\mathbb{P}}$ 的关系由 (11) 式决定, 那么随机变量 X 关于 $\tilde{\mathbb{P}}$ 的期望为

$$\tilde{\mathbb{E}}[X]=\mathbb{E}[XL] \tag{13}$$

为验证 (13) 式, 注意到对示性函数, (13) 式是成立的。对示性函数的线性组合取极限可得对一般的 X, (13) 式也是成立的。

性质 2 在测度 $\tilde{\mathbb{P}}$ 下, 条件期望有如下表示

$$\tilde{\mathbb{E}}[X|\mathcal{G}]=\frac{\mathbb{E}[XL|\mathcal{G}]}{\mathbb{E}[L|\mathcal{G}]} \tag{14}$$

现说明 (14) 式是成立的。首先, 由定义可得, 对所有的 $Y\in\mathcal{G}$, 下式成立

$$\tilde{\mathbb{E}}\{\tilde{\mathbb{E}}[X|\mathcal{G}]Y\}=\tilde{\mathbb{E}}[XY] \tag{15}$$

按定义, (15) 式左边是

$$\mathbb{E}\{\tilde{\mathbb{E}}[X|\mathcal{G}]YI\} = \mathbb{E}\{\tilde{\mathbb{E}}[X|\mathcal{G}]\mathbb{E}[L|\mathcal{G}]Y\}$$

在上述推导中应用了关于条件期望的塔式性质。另外, (15) 式右边是

$$\mathbb{E}[XYL] = \mathbb{E}[\mathbb{E}[XL|\mathcal{G}]Y]$$

在上述推导中应用了关于条件期望的迭代性质。

因此, 由 (15) 式对所有的 $Y \in \mathcal{G}$ 成立, 可推出

$$\tilde{\mathbb{E}}[X|\mathcal{G}]\mathbb{E}[L|\mathcal{G}] = \mathbb{E}[XL|\ ]$$

上式就是 (14) 式的变形。

性质 3 对 $X \in \mathcal{G}$, 可得

$$\tilde{\mathbb{E}}_{\mathcal{G}}[X] = \tilde{\mathbb{E}}[X] = \mathbb{E}[XL] = \mathbb{E}[X\mathbb{E}[L|\mathcal{G}]] = \mathbb{E}_{\mathcal{G}}[X\mathbb{E}[L|\mathcal{G}]] \tag{16}$$

由 (16) 式, 即得

$$\frac{\mathrm{d}\tilde{\mathbb{P}}_{\mathcal{G}}}{\mathrm{d}\mathbb{P}_{\mathcal{G}}} = \mathbb{E}[L|\mathcal{G}] \tag{17}$$

性质 4 设在概率空间上存在滤波 $\mathbf{F} = \{\mathcal{F}_t\}_{0\leqslant t\leqslant T}$。简记 $\mathbb{P}_t = \mathbb{P}_{\mathcal{F}_t}$。引入

$$L_t = \frac{\mathrm{d}\tilde{\mathbb{P}}_t}{\mathrm{d}\mathbb{P}_t}$$

在 (17) 式中取 $\mathcal{G} = \mathcal{F} = \mathcal{F}_T$, 可得 $L_T = L$。由 (17) 式可得

$$L_t = \mathbb{E}[L|\mathcal{F}_t] \tag{18}$$

(18) 式表明 $(L_t)_{0\leqslant t\leqslant T}$ 是一个 $(\mathbf{F},\mathbb{P})$ 鞅。

3. Girsanov 定理

Girsanov 定理是随机过程理论的著名结论之一, 它是数理金融中无套利定价理论的基础。下面在计数过程的框架下给出该定理。

定理 2 设 N_t 是计数过程, 它的 $(\mathbf{F},\mathbb{P})$ 强度为 λ_t。设 $\tilde{\lambda}_t$ 是非负的 $\mathbf{F}$ 可适应过程, 而且 $\tilde{\lambda}_t = 0$ 当且仅当 $\lambda_t = 0$。那么存在概率测度 $\tilde{\mathbb{P}}$, 使得 $\tilde{\mathbb{P}} \sim \mathbb{P}$, 而且 N 的 $(\mathbf{F},\tilde{\mathbb{P}})$ 强度为 $\tilde{\lambda}_t$。(18) 式的具体形式为

$$L_t = \exp\left(\int_0^t (\ln\tilde{\lambda}_s - \ln\lambda_s)\mathrm{d}N_s + \int_0^t (\lambda_s - \tilde{\lambda}_s)\mathrm{d}s\right)$$

证明　这里给出的证明是构造性的, 出发点是 (12) 式的 L。可考虑 $L = L_T$, 其中

$$L_t = \exp\left(\int_0^t \phi_s \mathrm{d}N_s + \int_0^t \psi_s \mathrm{d}s\right)$$

其中 ϕ 是可预料的, ψ 是可适应的。

首先要求 L_t 是 $(\mathbf{F},\mathbb{P})$ 鞅。由 Itô 公式,

$$\begin{aligned}\mathrm{d}L_t &= L_t\psi_t\mathrm{d}t + L_{t-}(\mathrm{e}^{\phi_t}-1)\mathrm{d}N_t\\ &= L_t(\psi_t + (\mathrm{e}^{\phi_t}-1)\lambda_t)\mathrm{d}t + L_{t-}(\mathrm{e}^{\phi_t}-1)\mathrm{d}M_t\end{aligned}$$

由 §4 中的 (9) 式, 为使 L_t 是鞅, 须要求

$$\psi_t = (1-\mathrm{e}^{\phi_t})\lambda_t \tag{19}$$

此时

$$\mathrm{d}L_t = L_{t-}(\mathrm{e}^{\phi_t}-1)\mathrm{d}M_t$$

其次须确定 ϕ_t, 使得下面的过程

$$\mathrm{d}\tilde{M}_t = \mathrm{d}N_t - \tilde{\lambda}_t\mathrm{d}t \tag{20}$$

是 $(\mathbf{F},\tilde{\mathbb{P}})$ 鞅, 即要求 $\tilde{\mathbb{E}}[\tilde{M}_t|\mathcal{F}_s] = \tilde{M}_s$。由(14)式, 即要求

$$\frac{\mathbb{E}[\tilde{M}_tL|\mathcal{F}_s]}{\mathbb{E}[L|\mathcal{F}_s]} = \tilde{M}_s$$

由(18)式和 §2 节的(5)式, 上式就是

$$\mathbb{E}[\tilde{M}_tL_t|\mathcal{F}_s] = \tilde{M}_sL_s$$

因此 $\tilde{M}_tL_t$ 是 $(\mathbf{F},\mathbb{P})$ 鞅。因为

$$\begin{aligned}\mathrm{d}(\tilde{M}_tL_t) &= (-\tilde{\lambda}_t\mathrm{d}t)L_t + \tilde{M}_t(\mathrm{e}^{\phi_t}-1)L_t(-\lambda_t\mathrm{d}t)\\ &\quad+\left((\tilde{M}_{t-}+1)L_{t-}\mathrm{e}^{\phi_t} - \tilde{M}_{t-}L_{t-}\right)\mathrm{d}N_t\\ &= L_t\mathrm{d}t\left(-\tilde{\lambda}_t + \mathrm{e}^{\phi_t}\lambda_t\right) + \left((\tilde{M}_{t-}+1)L_{t-}\mathrm{e}^{\phi_t} - \tilde{M}_{t-}L_{t-}\right)\mathrm{d}M_t\end{aligned}$$

因此, 为使 $\tilde{M}_tL_t$ 是 $(\mathbf{F},\mathbb{P})$ 鞅, 须要求 $\phi_t = \ln\tilde{\lambda}_t - \ln\lambda_t$。

第五章　离散时间鞅

§1　条件期望

1. 概率空间与变量

以下考虑离散时间的取实数值的随机过程 $X=\{X_1,X_2,\cdots\}$。设随机过程 X 定义于某概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上。在任意时刻 n, 可以观察到 $X_1,X_2,\cdots,X_n$ 的值, 而不能观察到 $X_{n+1},X_{n+2},\cdots$ 的值, 但已有的观察值一般会影响到未来的变化。所有可观察到的事件都包含于可测空间 $(\Omega,\mathcal{F})$ 中。为理解这一点, 考虑下面的例子。

例 1　(i) 掷骰子模型。设 $\Omega=\{1,2,3,4,5,6\}$, $\mathcal{F}$ 是 Ω 的所有子集构成的集合, $\mathbb{P}(\{i\})=\dfrac{1}{6}$ 对所有 i 成立。掷骰子一次, 当关心是否出现奇数时, 事件的集合可选为 $\mathcal{G}=\{\varnothing,\{1,3,5\},\{2,4,6\},\Omega\}$。

(ii) 一般地, 设 $X:\Omega\to\{1,2,\cdots,n\}$ 是随机变量,$\mathcal{F}$ 是 Ω 的所有子集构成的集合。当关心是否出现 $\{Y=k\}$ 时, 事件的集合可选为 $\mathcal{G}$, 其中 $\mathcal{G}=\sigma(Y)$。

(iii) 对随机过程 X, 记 $\mathcal{F}_n=\sigma(X_1,\cdots,X_n)$, 那么 $\mathcal{F}_n$ 表示到时刻 n 时所有的信息。显然 $\mathcal{F}_n\subset\mathcal{F}_{n+1}$, 即 $\mathcal{F}_{n+1}$ 与 $\mathcal{F}_n$ 相比, 包含了更多的信息 (或事件)。另外, $\mathcal{F}_n$ 是最小的 σ- 代数, 使得 $X_1,X_2,\cdots,X_n$ 是可测的。

例 2　考虑标准的二叉树结构。设在时刻 0 时状态为 0, 到达时刻 1 时, 状态增加或减少 1；从时刻 1 的状态出发, 到达时刻 2 时, 状态也增加或减少 1, 依次类推。那么到达时刻 n, 可能的轨道有 2^n 条。但在时刻 0 时, 所有的轨道都是可能的, 所以无法区分轨道, 信息最少。而到时刻 1 时, 根据时刻 1 到达的状态, 已经把轨

道加以区分, 信息增加了。同样可考虑以后时刻的信息, 它们是逐渐增加的。

定义 1 称满足 $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ 的一系列 σ- 代数 $\{\mathcal{F}_n : n \geqslant 1\}$ 为滤波。

由例 1 的 (iii) 可知, 每一个随机过程可产生一个滤波 $\{\mathcal{F}_n : n \geqslant 1\}$。

2. 条件期望

显然, 新信息的出现会影响到概率的变化。现在考虑给定的信息如何影响事件的概率这一问题。例如, 在例 1 的 (i) 中, 如果已知骰子出现偶数, 那么出现 3 的条件概率就是 0。

例 3 设 X 和 Y 是定义于 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的两个不独立的随机变量, $X : \Omega \to \{1, 2, \cdots, n\}$, $Y : \Omega \to \{1, 2, \cdots, m\}$。现在考虑 Y 出现的结果如何影响到 X 的结果。首先考虑条件概率

$$\mathbb{P}[X = i | Y = j] = \frac{\mathbb{P}[X = i, Y = j]}{\mathbb{P}[Y = j]}$$

如果已经观察到事件 $\{Y = j\}$, 那么 X 可理解为定义于一个不同的概率空间上。在新的概率空间下, 概率测度定义为

$$\mathbb{P}[A | Y = j] = \frac{\mathbb{P}[A \cap \{Y = j\}]}{\mathbb{P}[\{Y = j\}]}$$

X 的条件期望为

$$\mathbb{E}[X | Y = j] = \sum_{k=1}^{n} k \mathbb{P}[X = k | Y = j]$$

上述条件期望是一个随机变量,

$$\mathbb{E}[X | Y] : \Omega \to \mathbb{R}, \mathbb{E}[X | Y](\omega) = \mathbb{E}[X | Y = Y(\omega)]$$

条件期望 $\mathbb{E}[X|Y]$ 有以下性质:

(i) 它是 $\sigma(Y)$ 可测的。

(ii) 对每个 $G \in \sigma(Y)$, 下列关系成立

$$\int_G \mathbb{E}[X|Y](\omega) \mathrm{d}\mathbb{P}(\omega) = \int_G X(\omega) \mathrm{d}\mathbb{P}(\omega)$$

性质 (i) 表明 $\mathbb{E}[X|Y]$ 在每个集合 $\{Y=j\}$ 上是常数。为说明性质 (ii), 注意到任意 $G\in\sigma(Y)$ 可表示为 $Y^{-1}(A)$, 其中 $A\in\{1,2,\cdots,m\}$。因此

$$\begin{aligned}
&\int_G \mathbb{E}[X|Y](\omega)\mathrm{d}\mathbb{P}(\omega)\\
&=\int_{Y^{-1}(A)}\mathbb{E}[X|Y](\omega)\mathrm{d}\mathbb{P}(\omega)=\sum_{j\in A}\mathbb{E}[X|Y=j]\mathbb{P}[Y=j]\\
&=\sum_{j\in A}\sum_{k=1}^{n}k\mathbb{P}[X=k|Y=j]\mathbb{P}[Y=j]=\sum_{j\in A}\sum_{k=1}^{n}k\mathbb{P}[X=k,Y=j]\\
&=\int\sum_{j\in A}\sum_{k=1}^{n}k1_{\{X=k\}}1_{\{Y=j\}}\mathrm{d}\mathbb{P}(\omega)=\int X(\omega)\sum_{k=1}^{n}1_{\{X=k\}}\sum_{j\in A}1_{\{Y=j\}}\mathrm{d}\mathbb{P}(\omega)\\
&=\int X(\omega)\sum_{j\in A}1_{\{Y=j\}}\mathrm{d}\mathbb{P}(\omega)=\int_{Y^{-1}(A)}X(\omega)\mathrm{d}\mathbb{P}(\omega)\\
&=\int_G X(\omega)\mathrm{d}\mathbb{P}(\omega)
\end{aligned}$$

在一些具体的情形中, 可以直接求出条件期望。例如, 在投币试验中, 记 X_n 表示前 n 次试验中出现正面的次数, 那么容易得到

$$\mathbb{E}[X_{n+1}|X_n]=X_n+\frac{1}{2}$$

例 4 设 X 是参数为 λ 的 Poisson 变量, 它表示试验次数。如果 $X=n$, 假设每次试验出现成功的概率为 p, 而且不同试验之间是相互独立的。记 S 为总的成功次数, 求 $\mathbb{E}[S|X]$ 与 $\mathbb{E}[X|S]$。

首先计算 $\mathbb{E}[S|X=n]$。如果 $X=n$, 那么 S 是参数为 n 和 p 的二项分布变量, 因此 $\mathbb{E}[S|X=n]=np$, $\mathbb{E}[S|X]=pX$。

为计算 $\mathbb{E}[X|S]$, 首先计算 $\mathbb{E}[X|S=k]$, 为此对 $n\geqslant k$, 计算 $\mathbb{P}[X=n|S=k]$

$$\begin{aligned}
\mathbb{P}[X=n|S=k]&=\frac{\mathbb{P}[S=k|X=n]\mathbb{P}[X=n]}{\mathbb{P}[S=k]}\\
&=\frac{\binom{n}{k}p^k(1-p)^{n-k}\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}}{\sum\limits_{m\geqslant k}\binom{m}{k}p^k(1-p)^{m-k}\frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}}\\
&=\frac{[(1-p)\lambda]^{n-k}}{(n-k)!}\mathrm{e}^{-(1-p)\lambda}
\end{aligned}$$

因此

$$\mathbb{E}[X|S=k]=\sum_{n\geqslant k}n\frac{[(1-p)\lambda]^{n-k}}{(n-k)!}\mathrm{e}^{-(1-p)\lambda}=k+(1-p)\lambda$$
$$\mathbb{E}[X|S]=S+(1-p)\lambda$$

上述结论的直观解释是：试验次数等于成功次数与失败次数之和。

下面给出条件期望的一般性概念，至于条件期望的存在性证明，这里不再给出。

定义 2　设 X 是定义于概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量，$\mathbb{E}[|X|]<\infty$。$\mathcal{G}$ 是 $\mathcal{F}$ 的子 σ– 代数，$\mathcal{G}\subset\mathcal{F}$。$X$ 关于 $\mathcal{G}$ 的条件期望记为 $\mathbb{E}[X|\mathcal{G}]$，它满足以下两个条件：

(i) $\mathbb{E}[X|\mathcal{G}]$ 是 $\mathcal{G}$ 可测的。

(ii) 对所有的 $G\in\mathcal{G}$，下式成立

$$\int_G\mathbb{E}[X|\mathcal{G}]\mathrm{d}\mathbb{P}=\int_G X\mathrm{d}\mathbb{P}$$

直观上看，由于 $\mathcal{G}$ 包含的集合较少，关于 $\mathcal{G}$ 可测的变量相对于关于 $\mathcal{F}$ 可测的变量，取值要少些。上述条件 (i) 表明，条件期望变量 $\mathbb{E}[X|\mathcal{G}]$ 是原来变量 X 的光滑化。在上述条件 (ii) 中取 $G=\Omega$，即得 $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]]=\mathbb{E}[X]$。下面给出条件期望的一些性质。

(i) 对实数 a 和 b，随机变量 X 和 Y，

$$\mathbb{E}[aX+bY|\mathcal{G}]=a\mathbb{E}[X|\mathcal{G}]+b\mathbb{E}[Y|\mathcal{G}]$$

(ii) 如果 $X\geqslant 0$，那么 $\mathbb{E}[X|\mathcal{G}]\geqslant 0$。

(iii) 如果 $\mathcal{G}\subset\mathcal{H}$，那么 $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]=\mathbb{E}[X|\mathcal{G}]$。

(iv) 如果 $Y\in\mathcal{G}$，那么 $\mathbb{E}[XY|\mathcal{G}]=Y\mathbb{E}[X|\mathcal{G}]$。特别地，如果 $X\in\mathcal{G}$，那么 $\mathbb{E}[X|\mathcal{G}]=X$。

(v) 如果 X 与 $\mathcal{F}$ 独立，那么 $\mathbb{E}[X|\mathcal{G}]=\mathbb{E}[X]$。

下面仅证明性质 (iii)。首先 $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]\in\mathcal{G}$，对任意事件 $G\in\mathcal{G}\subset\mathcal{H}$，应用两次条件期望的定义，即得

$$\int_G\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]\mathrm{d}\mathbb{P}=\int_G\mathbb{E}[X|\mathcal{H}]\mathrm{d}\mathbb{P}=\int_G X\mathrm{d}\mathbb{P}=\int_G\mathbb{E}[X|\mathcal{G}]\mathrm{d}\mathbb{P}$$

上式就表明

$$\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}]=\mathbb{E}[X|\mathcal{G}]$$

§2 鞅与下鞅

1. 定义与例子

鞅这一术语最初表示公平赌博策略。

定义 3 设 $\mathbf{F} = \{\mathcal{F}(n) : n \geqslant 0\}$ 是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的滤波, $X = \{X_n : n \geqslant 0\}$ 是离散时间随机过程。如果下列条件满足

(i) $X_n \in \mathcal{F}(n)$, 即 X 是可适应的过程。

(ii) $\mathbb{E}[|X_n|] < \infty$ 对所有的 n 成立。

(iii) $\mathbb{E}[X_n|\mathcal{F}(n-1)] = X_{n-1}$。

那么称 $\{X, \mathbf{F}\}$ 为鞅。

如果 (iii) 改为 $\mathbb{E}[X_n|\mathcal{F}(n-1)] \geqslant X_{n-1}$, 那么称 $\{X, \mathbf{F}\}$ 为下鞅。如果 (iii) 改为 $\mathbb{E}[X_n|\mathcal{F}(n-1)] \leqslant X_{n-1}$, 那么称 $\{X, \mathbf{F}\}$ 为上鞅。注意到在以上定义中, $\mathbf{F}$ 是预先给定的。在应用中, $\mathbf{F}$ 通常取为由 X 产生的滤波, 称为自然滤波。

对鞅 $\{X_n\}$,$m < n$, 由条件期望的性质可推出下式成立

$$\begin{aligned}\mathbb{E}[X_n|\mathcal{F}(m)] &= \mathbb{E}\{\mathbb{E}[X_n|\mathcal{F}(n-1)]|\mathcal{F}(m)\} = \mathbb{E}[X_{n-1}|\mathcal{F}(m)] \\ &= \cdots = \mathbb{E}[X_{m+1}|\mathcal{F}(m)] = X_m\end{aligned}$$

对上式两边取期望, 可推出对所有的 n, 下式成立

$$\mathbb{E}[X_n] = \mathbb{E}[X_m] = \cdots = \mathbb{E}[X_0]$$

另外由鞅的定义可得

$$\mathbb{E}[X_n - X_{n-1}|\mathcal{F}(n-1)] = 0$$

下面给出鞅的一些例子。

例 5 设 $X_1, X_2, \cdots$ 是独立随机变量, $\mathbb{E}[|X_n|] < \infty$ 而且 $\mathbb{E}[X_n] = 0$。记 $S_n = \sum_{k=1}^{n} X_k, \{\mathcal{F}(n) : n \geqslant 1\}$ 为随机过程 $\{S_n : n \geqslant 1\}$ 产生的滤波, 那么由条件期望的性质,

$$\mathbb{E}[S_n|\mathcal{F}(n-1)] = \mathbb{E}[S_{n-1} + X_n|\mathcal{F}(n-1)] = S_{n-1} + \mathbb{E}[X_n] = S_{n-1}$$

因此 $\{S_n : n \geqslant 1\}$ 是鞅。

例 6 设 $X_1, X_2, \cdots$ 是独立非负随机变量, $\mathbb{E}[X_n] = 1$。记 $M_n = \prod_{k=1}^{n} X_k, \{\mathcal{F}(n) : n \geqslant 1\}$ 为随机过程 $\{M_n : n \geqslant 1\}$ 产生的滤波, 那么

$$\mathbb{E}[M_n|\mathcal{F}(n-1)] = M_{n-1}\mathbb{E}[X_n|\mathcal{F}(n-1)] = M_{n-1}\mathbb{E}[X_n] = M_{n-1}$$

因此 $\{M_n : n \geqslant 1\}$ 是鞅。

例 7 设 X 为随机变量, $\{\mathcal{F}(n) : n \geqslant 1\}$ 为任意的滤波。定义 $X_n = \mathbb{E}[X|\mathcal{F}(n)]$, X_n 可解释为到时刻 n 时关于 X 的累积数据。由条件期望的性质可以推出 $\{X_n : n \geqslant 1\}$ 是鞅, 即

$$\mathbb{E}[X_n|\mathcal{F}(n-1)] = \mathbb{E}\{\mathbb{E}[X|\mathcal{F}(n)]|\mathcal{F}(n-1)\} = \mathbb{E}[X|\mathcal{F}(n-1)] = X_{n-1}$$

例 8 Galton-Watson 过程。设 $\{p_0, p_1, p_2, \cdots\}$ 是取值于非负整数的后代个数概率分布, 该分布的期望值 $\mu = \sum\limits_{k=0}^{\infty} kp_k < \infty$。$\{Y_k^n, k, n \geqslant 1\}$ 是独立同分布随机变量, $\mathbb{P}\{Y_k^n = i\} = p_i$。定义 $X_0 = 1$,

$$X_{n+1} = \sum_{k=1}^{X_n} Y_k^{n+1}, n \geqslant 0$$

另外, 设 $\mathcal{F}(n) = \sigma(Y_k^m : m \leqslant n, k \geqslant 1)$, 那么 $\{X_n\}$ 关于滤波 $\mathcal{F}(n)$ 是可适应的, 而且

$$\mathbb{E}[X_{n+1}|\mathcal{F}(n)] = \sum_{k=1}^{\infty} \mathbb{E}\{1_{\{k \leqslant X_n\}} Y_k^{n+1}|\mathcal{F}(n)\} = \sum_{k=1}^{\infty} 1_{\{k \leqslant X_n\}}\mathbb{E}\{Y_k^{n+1}\} = \mu X_n$$

因此当 $\mu = 1$ 时 $\{X_n\}$ 是鞅, 当 $\mu \geqslant 1$ 时 $\{X_n\}$ 是下鞅。

例 9 设在时刻 0 时, 盒中有一白一黑两球。在每一时刻 $n = 1, 2, \cdots$, 随机地从盒中取一球再放回, 并额外放入盒中与抽出的球同样颜色的一球。这样到时刻 n 后, 盒中共有 $n+2$ 个球, 其中的黑球数为 $B_n + 1$, B_n 表示到时刻 n 时抽取的黑球次数。因此到时刻 n 后, 盒中黑球比例为

$$M_n = \frac{B_n + 1}{n+2}$$

记 $\{\mathcal{F}(n)\}$ 为 $\{B_n\}$ 产生的滤波, 下面证明 $\{M_n\}$ 是鞅。

首先, $B_n = X_1 + X_2 + \cdots + X_n$, 其中 X_n 表示在时刻 n 时抽取的黑球个数, 它取 0 或 1, 而且

$$\mathbb{P}[X_n = 1|B_{n-1} = k] = \frac{k+1}{n+1}, 0 \leqslant k \leqslant n-1, n \geqslant 2$$

当 $n \geqslant 1$ 时,

$$\mathbb{E}[M_n|\mathcal{F}(n-1)] = \mathbb{E}\left[\frac{B_{n-1}+X_n+1}{n+2}\Big|\mathcal{F}(n-1)\right] = \frac{B_{n-1}+1}{n+2} + \mathbb{E}\left[\frac{X_n}{n+2}\Big|\mathcal{F}(n-1)\right]$$
$$= \frac{B_{n-1}+1}{n+2} + \frac{B_{n-1}+1}{(n+1)(n+2)} = \frac{B_{n-1}+1}{n+1} = M_{n-1}$$

2. 鞅变换

设 $\{C_n : n \geqslant 1\}$ 关于 $\{\mathcal{F}(n)\}$ 是可预期过程, 即 $C_n \in \mathcal{F}(n-1)$。由鞅 $\{X_n : n \geqslant 0\}$ 与可预期过程 $\{C_n : n \geqslant 1\}$, 定义

$$Y_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1})$$

称过程 $\{Y_n : n \geqslant 1\}$ 为鞅 $\{X_n : n \geqslant 0\}$ 关于 $\{C_n : n \geqslant 1\}$ 的鞅变换。如果把鞅 $\{X_n : n \geqslant 0\}$ 视为公平赌博, C_n 表示在时刻 n 的赌注, 它应当在时刻 $n-1$ 决定, 那么 $C_n(X_n - X_{n-1})$ 表示第 n 次赌博的收益, 而 Y_n 表示前 n 次赌博的收益。下面的定理表明, 一般来说, 在公平赌博中, 净收益的期望值为 0。

定理 1 设 $\{X_n : n \geqslant 0\}$ 是鞅, 可预期过程 $\{C_n : n \geqslant 1\}$ 是有界的, 那么 $\{Y_n : n \geqslant 1\}$ 也是鞅, 而且 $\mathbb{E}[Y_n] = 0$。

证明 首先由定义,$Y_n \in \mathcal{F}(n)$。其次,

$$\mathbb{E}[Y_n|\mathcal{F}(n-1)] = \mathbb{E}\left\{\sum_{k=1}^{n} C_k(X_k - X_{k-1})|\mathcal{F}(n-1)\right\}$$
$$= \sum_{k=1}^{n-1} C_k(X_k - X_{k-1}) + C_n\mathbb{E}[X_n - X_{n-1}|\mathcal{F}(n-1)]$$
$$= \sum_{k=1}^{n-1} C_k(X_k - X_{k-1}) = Y_{n-1}$$

这就证明了 $\{Y_n : n \geqslant 1\}$ 是鞅。只需注意到 $\mathbb{E}[Y_1] = 0$, 从而 $\mathbb{E}[Y_n] = 0$。

3. Doob 可选停时定理

设 $T(\omega)$ 为取非负整数值的随机变量, 如果对任意 n,

$$\{\omega : T(\omega) \leqslant n\} \in \mathcal{F}(n)$$

那么称 T 为停时。

把鞅 $\{X_n\}$ 视为公平游戏, T 视为停止游戏规则, 它应由已有信息确定。进一步设在时刻 T 前的赌注都是一个单位, 即 $C_n = 1_{\{n \leqslant T\}}$, 注意到

$$C_n = 1_{\{n \leqslant T\}} = 1 - 1_{\{T \leqslant n-1\}}$$

从而 $C_n \in \mathcal{F}(n-1)$。由定理 1, 鞅变换后得到一个鞅。另一方面,

$$\sum_{k=1}^{n} C_k(X_k - X_{k-1}) = X_{T\wedge n} - X_0$$

其中 $x \wedge y = \min\{x, y\}, X_{T\wedge n} = X_{T(\omega)\wedge n}(\omega)$。总之, 即得下述定理。

定理 2 如果 X 是鞅, T 是停时, 那么 $X_{T\wedge n}$ 是鞅。特别地, 对所有的 n, 下式成立

$$\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$$

例 10 考虑对称随机游动。设 $Y_1, Y_2, \cdots$ 为一列独立同分布随机变量,$\mathbb{P}\{Y_n = 1\} = \mathbb{P}\{Y_n = -1\} = \dfrac{1}{2}$, 记 $X_n = \sum_{k=1}^{n} Y_k$, 那么 $\{X_n\}$ 是鞅。设

$$T = \inf\{n \geqslant 1 : X_n = 1\}$$

可以证明 T 是一个停时。定理 2 表明, 对所有的 n,$\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_1] = 0$。

显然在此例中, $X_T \equiv 1$, 从而 $\mathbb{E}[X_T] = 1$。因此, 一般地 $\lim\limits_{n\to\infty} \mathbb{E}[X_{T\wedge n}] \neq \mathbb{E}[X_T]$, 即极限与积分顺序不能交换。但在一定条件下, 极限与积分顺序可以交换。为此下面不加证明地给出 Doob 可选停时定理。

定理 3 (Doob 可选停时定理) 设 X 是鞅, T 是停时。如果以下条件之一成立, 那么

$$\mathbb{E}[X_T] = \mathbb{E}[X_0]$$

(i) T 是有界的。

(ii) T 是有限的, 而且 X 是有界的。

(iii) $\mathbb{E}[T] < \infty$, 而且存在 $K > 0$, 使得 $|X_n(\omega) - X_{n-1}(\omega)| \leqslant K$。

(iv) T 是有限的, 而且 X 是非负的。

4. Doob 可选停时定理的一个应用

设 $X_1, X_2, \cdots$ 为一列独立同分布随机变量,$\mathbb{P}\{X_i = 1\} = p, \mathbb{P}\{X_i = -1\} = q$, 其中 $p + q = 1, p > q > 0$。设 $\mathcal{F}(n) = \sigma(X_1, \cdots, X_n), 0 < a < b$, a 和 b 是两个整数,$S_n = a + X_1 + \cdots + X_n$, S_n 可表示为在一个有利的赌博中, 初始资本为 a, 经过 n 步赌博以后的资本。定义停时 T 如下

$$T = \inf\{n \geqslant 1 : S_n = 0 \text{或} S_n = b\}$$

下面求 $\mathbb{P}\{S_T = 0\}$, $\mathbb{E}[T]$, $\mathbb{E}[S_T]$。

首先证明 $\mathbb{E}[T] < \infty$。对充分大的 k, 如果 $T = k$, 那么把 k 以前的整数分组, 相邻的 b 个整数作为一组, 即 $(nb+1, \cdots, nb+b), nb+b < k$, 每一组 $(X_{nb+1}, \cdots, X_{nb+b})$

至少有一个 $X_i = -1$, 其概率为 $1-p^b$, 从而 $\mathbb{P}[T=k] \leqslant (1-p^b)^{(k-b)/b}$。因此 $\mathbb{P}[T=k]$ 以指数形式递减。这就证明了 $\mathbb{E}[T] < \infty$。

可以验证如下定义的两个过程 $\{M_n\}$ 和 $\{N_n\}$ 是鞅。

$$M_n = \left(\frac{q}{p}\right)^{S_n}, N_n = S_n - n(p-q)$$

对 $\{M_n\}$, 应用定理 3(iii), 注意到 $|M_n - M_{n-1}| \leqslant 2$, 可得

$$\left(\frac{q}{p}\right)^a = \mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{P}[S_T = 0] + \left(\frac{q}{p}\right)^b \mathbb{P}[S_T = b]$$

另外,

$$\mathbb{P}[S_T = b] = 1 - \mathbb{P}[S_T = 0]$$

从而

$$\mathbb{P}[S_T = 0] = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}$$

$$\mathbb{E}[S_T] = b\mathbb{P}[S_T = b] = b\frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^b}$$

同样地, 对 $\{N_n\}$, 应用定理 3(iii), 注意到 $|N_n - N_{n-1}| \leqslant 1 + p + q$, 可得

$$a = \mathbb{E}[N_0] = \mathbb{E}[N_T] = \mathbb{E}[S_T - T(p-q)] = \mathbb{E}[S_T] - \mathbb{E}[T](p-q)$$

最后得到

$$\mathbb{E}[T] = \frac{1}{p-q}(\mathbb{E}[S_T] - a) = \frac{b}{p-q}\frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^b} - \frac{a}{p-q}$$

5. Doob 分解定理

对一个可适应的过程 X, 是否存在一个过程 A, 使得两者的差 $X - A$ 是一个鞅? 这就是下面的 Doob 分解定理。

定理 4 (Doob 分解定理) 设 $\{\mathcal{F}(n) : n \geqslant 0\}$ 是一个滤波, $\{X_n : n \geqslant 0\}$ 是一个可适应过程, $\mathbb{E}[|X_n|] < \infty$, 那么 $\{X_n : n \geqslant 0\}$ 可分解为

$$X_n = X_0 + M_n + A_n$$

使得 $\{M_n : n \geqslant 0\}$ 是鞅, $M_0 = 0; \{A_n : n \geqslant 0\}$ 是可预期过程, $A_0 = 0$。另外, 上述分解是唯一的。

进一步, 如果 X 是下鞅, 那么 $\{A_n : n \geqslant 0\}$ 是递增的。

证明 先证唯一性。由分解式可得

$$
\begin{aligned}
\mathbb{E}[X_n - X_{n-1}|\mathcal{F}(n-1)] &= \mathbb{E}[M_n - M_{n-1}|\mathcal{F}(n-1)] + \mathbb{E}[A_n - A_{n-1}|\mathcal{F}(n-1)] \\
&= A_n - A_{n-1}
\end{aligned}
$$

最后一式由 $\{A_n : n \geqslant 0\}$ 的可预期性得到。另外, 由上式得到

$$
A_n = \sum_{k=1}^{n} \mathbb{E}[X_k - X_{k-1}|\mathcal{F}(k-1)]
$$

这就证明了 $\{A_n : n \geqslant 0\}$ 和 $\{M_n : n \geqslant 0\}$ 的唯一性。由 A 的构造可知, X 是下鞅等价于 A 是递增过程。

上述构造性方法, 也证明了分解的存在性。

现在考虑一个特例, 设 $\{M_n : n \geqslant 0\}$ 是鞅, $M_0 = 0, \mathbb{E}[M_n^2] < \infty$, 那么可以验证 $\{M_n^2 : n \geqslant 0\}$ 是下鞅。由定理 4, $\{M_n^2 : n \geqslant 0\}$ 有如下分解

$$
M_n^2 = N_n + A_n
$$

有时, 记 $A = \langle M \rangle$。因此, 对鞅 M, 存在可预期过程 $\langle M \rangle$, 使得 $M^2 - \langle M \rangle$ 是鞅。

§3 逆向随机游动

1. 逆向随机游动

设 $X_1, X_2, \cdots$ 是一列独立同分布的随机变量, $\mathbb{E}|X_1| < \infty$, 记 $S_n = X_1 + \cdots + X_n$。定义过程 $M_m = \dfrac{S_{n-m}}{n-m}, 0 \leqslant m \leqslant n-1$ 即考虑

$$
\frac{S_n}{n}, \frac{S_{n-1}}{n-1}, \cdots, \frac{S_2}{2}, \frac{S_1}{1}
$$

下面要证明: 对给定的 $n, \{M_m, 0 \leqslant m \leqslant n-1\}$ 是一个鞅。

首先, 对 $i < j \leqslant k$, 由对称性可得 $\mathbb{E}[X_i|S_k] = \mathbb{E}[X_j|S_k]$。对 j 从 1 到 k 求和, 可得

$$
k\mathbb{E}[X_i|S_k] = \sum_{j=1}^{k} \mathbb{E}[X_j|S_k] = \mathbb{E}[S_k|S_k] = S_k
$$

即 $\mathbb{E}[X_i|S_k] = \dfrac{S_k}{k}$, 此结论是很直观的, 这是因为 S_k 包含了 k 项。

由上述结论, 可得

$$\mathbb{E}[S_{k-1}|S_k] = \mathbb{E}[S_k - X_k|S_k] = S_k\left(1-\frac{1}{k}\right) = \frac{S_k}{k}(k-1)$$

即 $\mathbb{E}\left[\frac{S_{k-1}}{k-1}|S_k\right] = \frac{S_k}{k}$。另外, 注意到随机变量 $\{S_{k+1}-S_k, S_{k+2}-S_k, \cdots, S_n-S_k\}$ 与 $\{X_1, \cdots, X_k\}$ 是相互独立的, S_{k-1} 只与 $\{X_1, \cdots, X_k\}$ 有关, 从而

$$\mathbb{E}\left[\frac{S_{k-1}}{k-1}|S_k, \cdots, S_n\right] = \frac{S_k}{k}$$

在上式中令 $k = n-m$, 上式就是

$$\mathbb{E}[M_{m+1}|M_m, \cdots, M_0] = M_m$$

2. 投票定理

有了逆向随机游动, 下面给出一个定理。

定理 5 (投票定理) 两个候选人甲和乙, 甲的得票数为 a, 乙的得票数为 b, 总票数为 $n = a+b$。设甲当选, 即 $a > b$, 那么在对选票计数时, 甲总是领先的概率为 $\frac{a-b}{n}$。

证明 设 $X_1, X_2, \cdots$ 为独立同分布随机变量, X_1 取值为 0 和 2, 概率各为 1/2, 记 $S_m = X_1 + \cdots + X_m$。当选票是选甲时, X 取值为 0; 当选票是选乙时, X 取值为 2。那么甲总是领先这一事件可表示为

$$G = \{S_j < j, j \geqslant 1\}$$

所求的概率是 $\mathbb{P}[G|S_n = 2b]$。

应用逆向鞅

$$M_0 = S_n/n, M_1 = S_{n-1}/(n-1), \cdots, M_{n-1} = S_1/1$$

由假设, $M_0 = 2b/n < 1$。设 $T = \min\{m : M_m = 1\}$, T 表示甲和乙得票数首次相等的时刻, 如果不存在这样的时刻, 取 $T = n-1$。在事件 G^c 上, $M_T = 1$。而在事件 G 上, 甲总是领先的, 从而 $T = n-1$, 此时 $M_T = S_1/1 = 0$。由可选停时定理可得

$$\frac{2b}{n} = M_0 = \mathbb{E}[M_T] = \mathbb{P}(G^c) \times 1 + \mathbb{P}(G) \times 0$$

从上式得到 $\mathbb{P}(G^c) = \frac{2b}{n} = \frac{2b}{a+b}$, 最后得到

$$\mathbb{P}(G) = 1 - \mathbb{P}(G^c) = \frac{a-b}{n}$$

第六章　连续时间鞅

§1　Brown运动与 Poisson 过程

早在 1900 年, Bachelier 完成了关于投机理论的论文, 研究了随机过程在金融市场中的应用。特别地, 股票市场指数波动与悬浮在液面上的粒子运动很相似。Brown运动作为描述粒子运动的随机过程, 后来由 Einstein 加以研究。因此, 一个世纪以前, 人们已经了解到了Brown运动与金融的联系。几乎同时, 在 1903 年, Lundberg 发表了奠基性的论文, 以随机过程的观点建立了非寿险精算的数学基础。

在以下讨论中, 设所讨论的随机过程定义在某概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上。

1. 基本过程

定义 1　如果随机过程 W 满足以下条件

(i) $W_0 = 0$。

(ii) W 具有独立增量性: 即对 $0 \leqslant t_0 < t_1 < \cdots < t_n < \infty$, 随机变量 $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \cdots, W_{t_n} - W_{t_{n-1}}$ 是相互独立的。

(iii) W 的增量服从正态分布: 对 $s < t, W_t - W_s \sim N(0, t-s)$。

(iv) W 的轨道是连续的。

那么称 W 为标准Brown运动。

上述条件 (ii) 和 (iii) 可概括为: W 具有平稳独立增量性, 而且增量服从正态分布。

由定义, 在给定 $W_t = x$ 时,W 的瞬时均值和方差分别为

$$\lim_{\Delta t \to 0} \frac{\mathbb{E}\Delta W_t}{\Delta t} = 0, \lim_{\Delta t \to 0} \frac{\mathrm{Var}\Delta W_t}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta t}{\Delta t} = 1$$

直观上讲, 从离散时间离散状态的随机游动出发, 适当调整时间与空间单位, 然后构造折线轨道, 最后取极限可得到Brown运动。但关于Brown运动的存在性证明, 却是一个困难的问题。有关证明最早由 N.Wiener 给出。

定义 2 如果计数过程 N 满足以下条件

(i) $N(0) = 0$。

(ii) N 具有平稳独立增量性。

(iii) 对 $0 \leqslant s < t < \infty, N(t) - N(s)$ 服从参数为 λ 的 Poisson 分布,

$$\mathbb{P}[N(t) - N(s) = k] = \mathrm{e}^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}, k = 0, 1, 2, \cdots \tag{1}$$

那么称 N 是强度参数为 λ 的齐次 Poisson 过程。

定义 3 设 N 为齐次 Poisson 过程, $\{X_i : i \geqslant 1\}$ 为一列独立同分布随机变量, 定义过程 S 如下

$$S(t) = \sum_{k=1}^{N(t)} X_k$$

称过程 S 为复合 Poisson 过程。

由全概率公式和独立性假设, 可以计算 $\mathbb{E}[S_t]$ 如下

$$\begin{aligned}\mathbb{E}[S_t] &= \sum_{n \geqslant 0} \mathbb{E}[S_t | N_t = n] \mathbb{P}[N_t = n] = \sum_{n \geqslant 0} n \mathbb{E}[X_1] \mathbb{P}[N_t = n] \\ &= \sum_{n \geqslant 0} \mathbb{E}[X_1] n \mathbb{P}[N_t = n] = \mathbb{E}[X_1] \mathbb{E}[N_t] = \lambda t \mathbb{E}[X_1]\end{aligned}$$

类似地可得到 $\mathrm{Var}[S_t] = \lambda t \mathbb{E}[X_i^2]$。

在保险应用中, $N(t)$ 表示在时间区间 $(0, t]$ 内来自一个保单组合中索赔发生的次数。记 S_n 为第 n 次索赔发生时刻, 那么

$$N(t) = \max\{n \geqslant 1 : S_n \leqslant t\}, \ t \geqslant 0$$

相临索赔的时间记为 T_k, 如记 $S_0 = 0$, 那么 $T_k = S_k - S_{k-1}, k = 1, 2, \cdots$。$\{T_k : k \geqslant 1\}$ 独立同分布, 而且有指数分布, 参数为 λ。索赔额变量 $\{X_k : k \geqslant 1\}$ 是独立同分布的, 分布函数记为 F, 假设 $F(0) = 0, \mathbb{E}(X) = \mu$。因此, 到时刻 t 时, 总索赔额为 $S(t) = \sum_{k=1}^{N(t)} X_k$, $\{S(t) : t \geqslant 0\}$ 是一个复合 Poisson 过程。另一方面, 记保费收取过

程为 $\{P(t): t \geqslant 0\}$, 此处暂时考虑确定性的线性过程 $P(t) = u + ct$, 其中 $u \geqslant 0$ 为初始资本。

至此, 定义风险过程为

$$U(t) = u + ct - S(t), t \geqslant 0 \tag{2}$$

风险过程 $\{U(t): t \geqslant 0\}$ 的破产时刻记为 τ

$$\tau = \inf\{t \geqslant 0 : U(t) < 0\} \tag{3}$$

风险过程 $\{U(t): t \geqslant 0\}$ 的破产概率记为

$$\psi(u, t) = \mathbb{P}[\tau \leqslant t], t \leqslant \infty \tag{4}$$

以 $\psi(u, \infty)$ 表示无限时间内的破产概率, 简记为 $\psi(u)$。对风险过程, 通常要求下列正附加性条件

$$c - \lambda\mu > 0 \tag{5}$$

(5) 式表明, 平均来说, 保费收入要大于索赔损失, 此时可以证明 $\lim\limits_{u\to\infty} \psi(u) = 0$。

定义 4　称满足 (5) 式的风险过程 $\{U(t): t \geqslant 0\}$ 为 Cramér-Lundberg 风险过程。

以下结论以多种形式出现于应用随机过程的文献中。

定理 1　Cramér-Lundberg 风险过程的破产概率有以下表示

$$1 - \psi(u) = (1-\rho)\sum_{n=0}^{\infty} \rho^n F_I^{n*}(u), \quad u \geqslant 0$$

其中 $\rho = \lambda\mu/c < 1$, 如记 $\overline{F} = 1 - F$, 那么

$$F_I(x) = \frac{1}{\mu}\int_0^x \overline{F}(y)\mathrm{d}y, \quad x \geqslant 0$$

对取值于 $[0, \infty)$ 的分布 H, H^{n*} 表示 H 的 n 重卷积。由于当 $n = 0$ 时对应于空和, 因此定义

$$H^{*0}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases}$$

定理 1 表明 $1 - \psi(u)$ 可视为一个复合几何分布的分布函数, 或等价地 $\psi(u)$ 可视为一个复合几何分布的尾概率。

2. 关于鞅的基本结论

由标准Brown运动和齐次 Poisson 过程, 可以构造以下鞅。

定理 2 设 W 为标准Brown运动, 那么

(i) W 与 $\{W_t^2 - t : t \geqslant 0\}$ 都是鞅。

(ii) 对任意 $\mu \in \mathbb{R}, \sigma > 0$, 记 $W_{\mu,\sigma}(t) = \mu t + \sigma W$, 称 $\{W_{\mu,\sigma}(t)\}_{t\geqslant 0}$ 为漂移项为 μ, 方差为 σ^2 的Brown运动。对每个 $\beta \in \mathbb{R}$, 定义

$$\varepsilon(W)_t = \exp\{\beta W_{\mu,\sigma}(t) - [\mu\beta + (\sigma^2\beta^2/2)]t\} \tag{6}$$

那么 $\{\varepsilon(W)_t : t \geqslant 0\}$ 是一个鞅, 称为指数鞅。

定理 3 设 N 为强度参数为 λ 的齐次 Poisson 过程, 那么 $\{N(t) - \lambda t : t \geqslant 0\}$ 是一个鞅。

定理 4 对 Cramér-Lundberg 风险过程 $\{U(t) : t \geqslant 0\}$, 对 $r \in \mathbb{R}$, 定义

$$\theta(r) = \lambda(\mathbb{E}[\mathrm{e}^{rX_1}] - 1) - cr \tag{7}$$

$$M_r(t) = \exp\{-rU(t) - \theta(r)t\} \tag{8}$$

那么 $\{M_r(t) : t \geqslant 0\}$ 是一个鞅。

证明 对 $0 \leqslant s < t$, 反复应用平稳独立增量性,

$$\begin{aligned}
\mathbb{E}[M_r(t)|\mathcal{F}_s] &= \mathbb{E}[\exp\{-rU(t) - \theta(r)t\}|\mathcal{F}_s] \\
&= \mathbb{E}[\exp\{-r(U(t) - U(s))\}\exp\{-rU(s)\}|\mathcal{F}_s]\mathrm{e}^{-\theta(r)t} \\
&= \mathbb{E}[\exp\{-r(U(t) - U(s))\}|\mathcal{F}_s]\exp\{-rU(s) - \theta(r)t\} \\
&= \mathbb{E}[\exp\{-r\sum_{i=N(s)+1}^{N(t)} X_i\}|\mathcal{F}_s]\exp\{-rU(s) - \lambda(\mathbb{E}[\mathrm{e}^{rX_1}] - 1)t + crs\} \\
&= \exp\{\lambda(\mathbb{E}[\mathrm{e}^{rX_1}] - 1)(t - s)\}\exp\{-rU(s) - \lambda(\mathbb{E}[\mathrm{e}^{rX_1}] - 1)t + crs\} \\
&= \exp\{-rU(s) - \theta(r)s\} \\
&= M_r(s)
\end{aligned}$$

实际上, 以上三个定理的证明本质上都用到平稳独立增量性。

§2　二次变差过程

1. Doob-Meyer 分解定理

设在概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上存在滤波 $\mathbf{F} = \{\mathcal{F}_t\}_{t\geqslant 0}$, 随机过程 $X = \{X_t\}_{t\geqslant 0}$ 关于 $\mathbf{F}$ 是可适应的。如果 X 是可积的 (即对所有的 t, $\mathbb{E}|X_t| < \infty$), 而且当 $s \leqslant t$ 时, 下式成立

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s$$

那么称 X 为鞅。一般地, 如果 $\mathbb{E}[X_t|\mathcal{F}_s] \geqslant X_s$, 那么称 X 为下鞅。

对满足一定条件的右连续下鞅 X, Doob-Meyer 分解定理给出了一个分解, 即

$$X = M + A$$

其中 $M = \{M_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ 是右连续鞅, $A = \{A_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ 是可预期的增过程, 而且分解是唯一的。称过程 A 是可预期的, 是指 A_t 可由 t 之前的严格过去历史来确定, 即 $A_t \in \mathcal{F}_{t-}$, 其中

$$\mathcal{F}_{t-} = \bigcup_{\varepsilon>0} \mathcal{F}_{t-\varepsilon}$$

这里不对 Doob-Meyer 分解定理的条件进行详细讨论。只需指出, 本书所讨论的在精算学中出现的随机过程都满足这些条件。

2. 连续平方可积鞅

设 X 是右连续鞅。如果对所有的 $t \geqslant 0$, $\mathbb{E}[X_t^2] < \infty$, 那么称 X 是平方可积鞅。如果 $X_0 = 0$, 那么记 $X \in \mathcal{M}_2$。进一步, 如果 X 是连续的, 那么记 $X \in \mathcal{M}_2^c$。

(i) 设 $X \in \mathcal{M}_2, 0 \leqslant s < t \leqslant u < v$, 那么

$$\mathbb{E}[(X_v - X_u)(X_t - X_s)] = \mathbb{E}\{\mathbb{E}[(X_v - X_u)|\mathcal{F}_u](X_t - X_s)\} = 0$$

即在不同区间上的增量的乘积, 它的期望值为 0。

(ii) 设 $X \in \mathcal{M}_2$, 那么 $X^2 = \{X_t^2, \mathcal{F}_t; 0 \leqslant t < \infty\}$ 是非负下鞅, 这由下面的 Jensen 不等式即证。

引理 1 (Jensen 不等式)　设 $\varphi : \mathbb{R} \to \mathbb{R}$ 是凸函数,X 和 $\varphi(X)$ 是可积随机变量, 那么对任意 σ- 代数 $\mathcal{G}$, 下列不等式成立

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leqslant \mathbb{E}[\varphi(X)|\mathcal{G}]$$

证明 引用实分析的结论, 如果 $\varphi:\mathbb{R}\to\mathbb{R}$ 是凸函数, 那么存在可列的实数对 (a_n,b_n), 使得 $\varphi(x)=\sup_n(a_nx+b_n)$。一方面

$$\mathbb{E}[a_nX+b_n|\mathcal{G}]\leqslant\mathbb{E}[\varphi(X)|\mathcal{G}]$$

另一方面, $\mathbb{E}[a_nX+b_n|\mathcal{G}]=a_n\mathbb{E}[X|\mathcal{G}]+b_n$, 从而

$$a_n\mathbb{E}[X|\mathcal{G}]+b_n\leqslant\mathbb{E}[\varphi(X)|\mathcal{G}]$$

对上式取极限, 即证 Jensen 不等式。

由 Doob-Meyer 分解定理, 可得

$$X_t^2=M_t+A_t$$

其中 M 是右连续鞅, A 是可预期的增过程。如取 $M_0=A_0=0$, 那么分解是唯一的。

定义 5 对 $X\in\mathcal{M}_2$, 称初值为零的 A 为 X 的二次变差过程, 记为 $\langle X\rangle$。

例 1 设 $\{N_t,\mathcal{F}_t;0\leqslant t<\infty\}$ 是 Poisson 过程, 那么容易验证 $M_t=N_t-\lambda t$ 是鞅, 而且 $M\in\mathcal{M}_2,\langle M\rangle_t=\lambda t$。

设过程 X 与 Y 都属于 $\mathcal{M}_2$, 那么由定义 5, $(X+Y)^2-\langle X+Y\rangle$ 与 $(X-Y)^2-\langle X-Y\rangle$ 都是鞅, 从而它们的差 $4XY-[\langle X+Y\rangle-\langle X-Y\rangle]$ 也是鞅。受此启发, 可得下一定义。

定义 6 设过程 X 与 Y 都属于 $\mathcal{M}_2$,X 和 Y 的交互变差过程记为

$$\langle X,Y\rangle=\frac{1}{4}[\langle X+Y\rangle-\langle X-Y\rangle]$$

上式表明 $\langle X,Y\rangle$ 是有界变差过程。

如果对所有的 $0\leqslant t<\infty$, $\langle X,Y\rangle_t=0$, 那么称 X与 Y 是正交的。

由定义 6 即知, $XY-\langle X,Y\rangle$ 是鞅。另外, 设 X 与 Y 都属于 $\mathcal{M}_2$, 那么容易推出, 对 $0\leqslant s<t<\infty$,

$$\begin{aligned}\mathbb{E}[(X_t-X_s)(Y_t-Y_s)|\mathcal{F}_s]&=\mathbb{E}[X_tY_t-X_sY_s|\mathcal{F}_s]\\&=\mathbb{E}[\langle X,Y\rangle_t-\langle X,Y\rangle_s|\mathcal{F}_s]\end{aligned}$$

因此, “X 与 Y 正交” 等价于 “在已知当前信息的条件下,X 与 Y 在 $[s,t]$ 上的增量是不相关的”。

3. 二次变差过程的另一种解释

设 $X = \{X_t; 0 \leqslant t < \infty\}$ 是一个过程, 对给定的 $t > 0$, 记 $\Pi = \{t_0, t_1, \cdots, t_m\}$ 为时间区间 $[0, t]$ 的一个分割, 即 $0 = t_0 < t_1 < \cdots < t_m = t, |\Pi| = \max\limits_{1 \leqslant k \leqslant m} |t_k - t_{k-1}|$

$$V_t^{(2)}(\Pi) = \sum_{k=1}^{m} |X_{t_k} - X_{t_{k-1}}|^2$$

当 $|\Pi| \to 0$ 时, 称 $V_t^{(2)}(\Pi)$ 的极限为 X 在 $[0, t]$ 上的二次变差。下面不加证明地给出一个定理, 它表明这里定义的二次变差和前面的二次变差是等价的。

定理 5 设 $X \in \mathcal{M}_2^c$, 那么当 $|\Pi| \to 0$ 时,$V_t^{(2)}(\Pi)$ 依概率收敛到 $\langle X \rangle_t$, 即对任给的 $\varepsilon > 0, \eta > 0$, 存在 $\delta > 0$, 使得当 $|\Pi| < \delta$ 时,

$$\mathbb{P}[|V_t^{(2)}(\Pi) - \langle X \rangle_t| > \varepsilon] < \eta$$

类似地, 对交互变差过程, 有如下定理。

定理 6 设 $X \in \mathcal{M}_2^c, Y \in \mathcal{M}_2^c, \Pi = \{t_0, t_1, \cdots, t_m\}$ 为时间区间 $[0, t]$ 的一个分割, 那么

$$\lim_{|\Pi| \to 0} \sum_{k=1}^{m} (X_{t_k} - X_{t_{k-1}})(Y_{t_k} - Y_{t_{k-1}}) = \langle X, Y \rangle_t$$

由定理 5 和定理 6 可得如下不等式

$$|\langle X, Y \rangle|^2 \leqslant \langle X \rangle \langle Y \rangle$$

例 2 设 $B = \{B_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ 是标准Brown运动, 由平稳独立增量性, 容易验证 $\{B_t^2 - t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ 是鞅, 因此 $\langle B \rangle_t = t$。进一步, 设 $\{(B_t^{(1)}, B_t^{(2)}, \cdots, B_t^{(d)}), \mathcal{F}_t; 0 \leqslant t < \infty\}$ 是 d 维标准Brown运动, 即每一个分量是标准Brown运动, 而且不同的分量是相互独立的, 那么交互变差过程为

$$\langle B^{(i)}, B^{(j)} \rangle_t = \delta_{ij} t, \ t \geqslant 0$$

§3 关于连续平方可积鞅的随机积分

1. 连续平方可积鞅的轨道

首先, 对一般的连续平方可积鞅 M, 如果 $\langle M \rangle_t$ 是严格单增的, 那么 M 的轨道在任意时间区间 $[a, b]$ 内不是有界变差的。这是因为, 如果 $\tilde{\Pi} = \{t_0, t_1, \cdots, t_m\}$ 是 $[a, b]$ 的一个划分, 由上一节的结论,

$$\lim_{|\tilde{\Pi}| \to 0} \sum_{k=1}^{m} |M_{t_k} - M_{t_{k-1}}|^2 = \langle M \rangle_b - \langle M \rangle_a > 0$$

假设 M 的轨道在 $[a,b]$ 内是有界变差的, 那么

$$\lim_{|\Pi|\to 0}\sum_{k=1}^{m}|M_{t_k}-M_{t_{k-1}}|<\infty$$

此时, 由于 M 的轨道是连续的,

$$\lim_{|\Pi|\to 0}\sum_{k=1}^{m}|M_{t_k}-M_{t_{k-1}}|^2\leqslant\left(\lim_{|\Pi|\to 0}\max_{k_i\in\Pi}|M_{t_k}-M_{t_{k-1}}|\right)\lim_{|\Pi|\to 0}\sum_{k=1}^{m}|M_{t_k}-M_{t_{k-1}}|=0$$

这与前面的假设是矛盾的, 所以 M 的轨道在任意时间区间 $[a,b]$ 内不是有界变差的。

由上述结论, 对一般的鞅 M, 不能按照轨道积分的概念定义下述积分

$$I_T(X)=\int_0^T X_t(\omega)\mathrm{d}M_t(\omega)$$

顺便指出, 应用前面的证明方法, 如果连续过程 X 的轨道是有界变差的, 那么容易验证 $\langle X\rangle\equiv 0$。

2. 简单过程关于鞅的随机积分

为介绍随机积分的概念, 先考虑简单过程。

定义 7 设随机过程 X 有如下形式

$$X_t(\omega)=\xi_0(\omega)1_{\{0\}}(t)+\sum_{i=0}^{\infty}\xi_i(\omega)1_{(t_i,t_{i+1}]}(t),0\leqslant t<\infty$$

其中 $0=t_0<t_1<\cdots,\ \lim\limits_{n\to\infty}t_n=\infty,\{\xi_n\}_{n\geqslant 0}$为随机变量, 而且 $\xi_n\in\mathcal{F}_{t_n},\sup\limits_{n\geqslant 0}|\xi_n(\omega)|\leqslant C<\infty$。称 X 为简单过程。

对简单过程 X, 可定义它关于鞅 M 的随机积分为如下形式的鞅变换。当 $t_n\leqslant t<t_{n+1}$ 时,

$$\begin{aligned}I_t(X)&=\int_0^t X_t\mathrm{d}M_t\\&=\sum_{i=0}^{n-1}\xi_i(M_{t_{i+1}}-M_{t_i})+\xi_n(M_t-M_{t_n})\\&=\sum_{i=0}^{\infty}\xi_i(M_{t\wedge t_{i+1}}-M_{t\wedge t_i})\end{aligned}$$

由定义可得 $I_t(X)$ 是连续的。下面给出 $I_t(X)$ 的一些性质, 其中前两个性质是显然的。

(i) $I_0(X) = 0$。

(ii) $I(\alpha X + \beta Y) = \alpha I(X) + \beta I(Y), \alpha, \beta \in \mathbb{R}$。

(iii) $\mathbb{E}[I_t(X)|\mathcal{F}_s] = I_s(X), s \leqslant t$, 即随机积分 $I(X)$ 是一个连续鞅。

为证明性质 (iii), 对 $0 \leqslant s < t < \infty$,分别对三种情形, 即 $s \leqslant t_i, t_i < s \leqslant t_{i+1}, s > t_{i+1}$, 证明下式成立

$$\mathbb{E}[\xi_i(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})|\mathcal{F}_s] = \xi_i(M_{s\wedge t_{i+1}} - M_{s\wedge t_i})$$

而上式可由假设 $\xi_i \in \mathcal{F}_{t_i}$, 以及 M 是鞅得到。

(iv) $\mathbb{E}[(I_t(X) - I_s(X))^2|\mathcal{F}_s] = \mathbb{E}\left[\int_s^t X_u^2 \mathrm{d}\langle M\rangle_u|\mathcal{F}_s\right]$, 即随机积分 $I(X)$ 是平方可积鞅, 而且 $I(X)$ 的二次变差过程为

$$\langle I(X)\rangle_t = \int_0^t X_u^2 \mathrm{d}\langle M\rangle_u$$

为证明性质 (iv), 设对给定的 $0 \leqslant s < t < \infty, t_{m-1} \leqslant s < t_m, t_n \leqslant s < t_{n+1}$,

$$\begin{aligned}
&\mathbb{E}[(I_t(X) - I_s(X))^2|\mathcal{F}_s] \\
&= \mathbb{E}\left[\left\{\xi_{m-1}(M_{t_m} - M_s) + \sum_{i=m}^{n-1}\xi_i(M_{t_{i+1}} - M_{t_i}) + \xi_n(M_t - M_{t_n})\right\}^2 \middle| \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\left\{\xi_{m-1}^2(M_{t_m} - M_s)^2 + \sum_{i=m}^{n-1}\xi_i^2(M_{t_{i+1}} - M_{t_i})^2 + \xi_n^2(M_t - M_{t_n})^2\right\}^2 \middle| \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\left\{\xi_{m-1}^2(\langle M\rangle_{t_m} - \langle M\rangle_s)^2 + \sum_{i=m}^{n-1}\xi_i^2(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i})^2 + \xi_n^2(\langle M\rangle_t - \langle M\rangle_{t_n})^2\right\}^2 \middle| \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\int_s^t X_u^2 \mathrm{d}\langle M\rangle_u \middle| \mathcal{F}_s\right]
\end{aligned}$$

在上面的推导中, 用到了平方可积鞅的性质, 即在不同区间上的增量的乘积, 它的期望值为 0。另外, 还用到了 $M^2 - \langle M\rangle$ 是鞅的结论。

(v) $\mathbb{E}[I_t(X)]^2 = \mathbb{E}\int_0^t X_u^2 \mathrm{d}\langle M\rangle_u$。

在性质 (iv) 中取 $s = 0$, 然后两边取期望, 即得性质 (v)。

3. 一般过程关于鞅的随机积分

一般的随机过程 X 关于鞅 M 的积分, 本质上是通过简单过程关于鞅 M 的积分来定义的, 它涉及鞅的极限定理。这里不再对一般理论展开, 只要求 X 是可预期过程, 而且 $M \in \mathcal{M}_2^c$。另外要求对所有的 $t > 0$,

$$\mathbb{E}\int_0^t X_u^2 \mathrm{d}\langle M\rangle_u < \infty$$

此时 X 关于鞅 M 的随机积分是唯一确定的连续平方可积鞅 $\{I_t(X),\mathcal{F}_t;0\leqslant t<\infty\}, I_t(X)=\int_0^t X_s\mathrm{d}M_s$, 而且 $\langle I(X)\rangle_t=\int_0^t X_u^2\mathrm{d}\langle M\rangle_u$。

上述随机积分具有前面介绍的简单过程关于鞅的随机积分的所有性质。另外，下面不加证明地给出如下性质。

(i) 记 X 和 Y 关于鞅 M 的随机积分分别为 $I(X)$ 和 $I(Y)$, 那么对 $0\leqslant s<t<\infty$,

$$\mathbb{E}[(I_t(X)-I_s(X))(I_t(Y)-I_s(Y))|\mathcal{F}_s]=\mathbb{E}\left[\int_s^t X_uY_u\mathrm{d}\langle M\rangle_u\middle|\mathcal{F}_s\right]$$

(ii) 记 X 和 Y 关于鞅 M 和 N 的随机积分分别为 $I^M(X)$ 和 $I^N(Y)$, 那么 $I^M(X)$ 和 $I^N(Y)$ 的二次变差过程为

$$\langle I^M(X),I^N(Y)\rangle_t=\int_0^t X_uY_u\mathrm{d}\langle M,N\rangle_u$$

(iii) 记 X 关于鞅 M 的随机积分为 $N=I^M(X)$,Y 关于 N 的随机积分为 $I^N(Y)$, 那么 $I^N(Y)=I^M(XY)$, 即 $\int_0^t Y_s\mathrm{d}N_s=\int_0^t X_sY_s\mathrm{d}M_s$。

最后不加证明地给出如下结论，它在金融数学的无套利定价理论中有重要应用。

定理 7 (鞅表示定理) 在一个概率空间中，设 M 是鞅，而且它的二次变差过程为正，那么对任意鞅 N, 它可表示为关于 M 的随机积分，即

$$N_t=N_0+\int_0^t\phi_s\mathrm{d}M_s$$

§4 Itô 公式与随机微分方程

1. Itô 公式

有了前面随机积分的定义，下面介绍 Itô 公式，为此先引入半鞅的定义。

定义 8 如果可适应的连续随机过程 $X=\{X_t,\mathcal{F}_t;0\leqslant t<\infty\}$ 有如下分解

$$X_t=X_0+M_t+B_t$$

其中 M 是连续鞅,B 是可适应的连续有界变差过程, $B_0=0$, 那么称 X 为半鞅。

Itô 公式表明，连续半鞅的光滑函数仍是连续半鞅，而且它给出了新的半鞅的分解。下面不加证明地给出两个定理。

定理 8　设 X 是有如上分解的连续半鞅，$f:\mathbb{R}\to\mathbb{R}$ 是二次可微函数，那么 $f(X_t)$ 也是半鞅，而且

$$f(X_t)=f(X_0)+\int_0^t f'(X_s)\mathrm{d}M_s+\int_0^t f'(X_s)\mathrm{d}B_s+\frac{1}{2}\int_0^t f''(X_s)\mathrm{d}\langle M\rangle_s$$

上式右边第二项是一个鞅，而第三项和第四项之和为有界变差过程。

有时候，Itô 公式也可写成如下微分形式

$$\begin{aligned}\mathrm{d}f(X_t)&=f'(X_t)\mathrm{d}M_t+f'(X_t)\mathrm{d}B_t+\frac{1}{2}f''(X_t)\mathrm{d}\langle M\rangle_t\\&=f'(X_t)\mathrm{d}X_t+\frac{1}{2}f''(X_t)\mathrm{d}\langle M\rangle_t\end{aligned}$$

下面定理给出了一般情形下的 Itô 公式。

定理 9　设 $\{M_t=(M_t^{(1)},\cdots,M_t^{(d)}),\mathcal{F}_t\}$ 是 d 维鞅，$\{B_t=(B_t^{(1)},\cdots,B_t^{(d)}),\mathcal{F}_t\}$ 是 d 维可适应的有界变差过程，$B_0=0$，记 $X_t=X_0+M_t+B_t$，其中 $X_0\in\mathcal{F}_0$。设多元函数 $f(t,x):[0,\infty)\times\mathbb{R}^d\to\mathbb{R}$ 关于 t 是一次可微的，关于 x 是二次可微的，那么

$$\begin{aligned}f(t,X_t)=&\ f(0,X_0)+\int_0^t\frac{\partial}{\partial t}f(s,X_s)\mathrm{d}s+\sum_{i=1}^d\int_0^t\frac{\partial}{\partial x_i}f(s,X_s)\mathrm{d}B_s^{(i)}\\&+\sum_{i=1}^d\int_0^t\frac{\partial}{\partial x_i}f(s,X_s)\mathrm{d}M_s^{(i)}\\&+\frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d\int_0^t\frac{\partial^2}{\partial x_i\partial x_j}f(s,X_s)\mathrm{d}\langle M^{(i)},M^{(j)}\rangle_s\end{aligned}$$

下面给出一些例子。

例 3　设 $f(x)=x^2$，对标准Brown运动 W 应用定理 8，即得

$$W_t^2=2\int_0^t W_s\mathrm{d}W_s+t$$

等价地

$$2\int_0^t W_s\mathrm{d}W_s=W_t^2-t$$

例 4　设 $f(t,x)=\exp(\sigma x+\mu t)$，记 $X_t=\exp(\sigma W_t+\mu t)$，对标准Brown运动 W 应用定理 9，即得

$$\mathrm{d}X_t=X_t\left(\sigma\mathrm{d}W_t+\left(\mu+\frac{1}{2}\sigma^2\right)\mathrm{d}t\right)$$

例 5 设 $f(t,x):[0,\infty)\times\mathbb{R}\to\mathbb{R}$ 关于 t 是一次可微的, 关于 x 是二次可微的, 记 $X_t=f(t,W_t)$, 对标准Brown运动 W 应用定理 9, 即得

$$\mathrm{d}X_t=\frac{\partial}{\partial t}f(t,X_t)\mathrm{d}t+\frac{\partial}{\partial x}f(t,X_t)\mathrm{d}W_t+\frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,X_t)\mathrm{d}t$$

为使 X_t 是鞅, 函数 $f(t,x)$ 应满足如下条件

$$\frac{\partial}{\partial t}f(t,x)+\frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,x)=0$$

特别地, 取 $f(t,x)=\exp\left(\alpha x-\frac{1}{2}\alpha^2 t\right)$, 那么 $\exp\left(\alpha W_t-\frac{1}{2}\alpha^2 t\right)$ 是鞅。

例 6 引入多项式函数

$$H_n(x,y)=\frac{\partial^n}{\partial\alpha^n}\exp\left(\alpha x-\frac{1}{2}\alpha^2 y\right)\bigg|_{\alpha=0};x,y\in\mathbb{R}$$

例如 $H_0(x,y)=1,H_1(x,y)=x,H_2(x,y)=x^2-y,H_3(x,y)=x^3-3xy,H_4(x,y)=x^4-6x^2y+3y^2$。

这些多项式满足以下递推关系

$$\frac{\partial}{\partial x}H_n(x,y)=nH_{n-1}(x,y);n=1,2,\cdots$$

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}H_n(x,y)+\frac{\partial}{\partial y}H_n(x,y)=0;n=0,1,2,\cdots$$

对标准Brown运动 W, 可以证明 $H_n(W_t,t)$ 是鞅。

顺便指出, 前面定义的 $H_n(x,y)$ 和 Hermite 多项式 $h_n(x)$ 有密切的关系

$$h_n(x)=\frac{(-1)^n}{\sqrt{n!}}\mathrm{e}^{x^2/2}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(\mathrm{e}^{-x^2/2})$$

$$H_n(x,y)=\sqrt{n!}y^{n/2}h_n(x/\sqrt{y})$$

2. 随机微分方程

例 4 的结论是一般的随机微分方程的特例。随机微分方程的一般形式是

$$\mathrm{d}X_t=b(t,X_t)\mathrm{d}t+\sigma(t,X_t)\mathrm{d}W_t$$

其中 W 是标准Brown运动,b 和 σ 是依赖于 t 和 X_t 的函数, 当它们满足一定条件时, 随机微分方程存在唯一解。

例 7 如下随机微分方程的解是存在的, 称为 CIR 随机利率过程, 它是金融经济学中的基本利率过程之一。

$$\mathrm{d}r_t = \kappa(\mu - r_t)\mathrm{d}t + \sigma\sqrt{r_t}\mathrm{d}W_t$$

其中 $\kappa, \mu, \sigma > 0, W$ 是标准Brown运动。

例 8 如下随机微分方程的解是存在的, 称为 Ornstein-Uhlenbeck(OU) 过程。

$$\mathrm{d}r_t = \kappa(\mu - r_t)\mathrm{d}t + \sigma\mathrm{d}W_t$$

其中 $\kappa, \mu, \sigma > 0, W$ 是标准Brown运动。

在金融经济学中, 例 7 和例 8 中的过程都表示随机利率过程, 它们都表明随机利率 r_t 趋近于长期利率 μ, 而趋近速度与二者的偏差成比例。在例 7 中, 波动性与当前利率有关, 而在例 8 中, 它与当前利率无关。

§5 测度变换与 Girsanov 定理

定义在概率空间 $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ 上的随机过程 X, 它的分布特征是与概率测度 $\mathbb{P}$ 相联系的。具体地讲, 设 W 关于 $\mathbb{P}$ 是标准Brown运动, $\mathbb{Q} \sim \mathbb{P}$, 那么一般地, W 关于 $\mathbb{Q}$ 不再是标准Brown运动。这一节讨论测度变换与过程分布之间的关系。

1. 连续时间过程的 Radon-Nikodym 导数

测度 $\mathbb{Q}$ 关于测度 $\mathbb{P}$ 的 Radon-Nikodym 导数记为 $\dfrac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$, 它是一个随机变量。对可适应的随机过程 X, 以下关系成立

(i) $\mathbb{E}_{\mathbb{Q}}(X_T) = \mathbb{E}_{\mathbb{P}}\left(\dfrac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}X_T\right)$。

(ii) $\mathbb{E}_{\mathbb{Q}}(X_t|\mathcal{F}_s) = \varsigma_s^{-1}\mathbb{E}_{\mathbb{P}}(\varsigma_t X_t|\mathcal{F}_s), s \leqslant t \leqslant T$, 其中 $\varsigma_t = \mathbb{E}_{\mathbb{P}}\left(\dfrac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\middle|\mathcal{F}_t\right)$ 为 Radon-Nikodym 导数过程。

2. 一个简单的测度变换

设 W 关于测度 $\mathbb{P}$ 是标准Brown运动, 给定正数 $T > 0$, 测度 $\mathbb{Q}$ 由下列导数定义

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp(-\gamma W_T - \frac{1}{2}\gamma^2 T)$$

那么 W 关于测度 $\mathbb{Q}$ 是具有常数漂移项 $-\gamma$ 的Brown运动, 即 $W_t = \tilde{W}_t - \gamma t$, 其中 $\tilde{W}_t$ 关于测度 $\mathbb{Q}$ 是标准Brown运动。

为证明上述结论, 首先要说明 W_T 关于测度 $\mathbb{Q}$ 是正态分布变量 $N(-\gamma T, T)$, 这是因为

$$\mathbb{E}_{\mathbb{Q}}[\exp(\theta W_T)] = \mathbb{E}_{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\exp(\theta W_T)\right] = \mathbb{E}_{\mathbb{P}}\left[\exp(-\gamma W_T - \frac{1}{2}\gamma^2 T + \theta W_T)\right]$$
$$= \exp\left[-\frac{1}{2}\gamma^2 T + \frac{1}{2}(\theta-\gamma)^2 T\right] = \exp(-\theta\gamma T + \frac{1}{2}\theta^2 T)$$

进一步, 为证明 $\tilde{W}$ $(\tilde{W}_t = W_t + \gamma t)$ 关于测度 $\mathbb{Q}$ 是标准Brown运动, 需证明

(i) $\tilde{W}$ 是连续的, 而且 $\tilde{W}_0 = 0$。这是显然的。

(ii) $\tilde{W}_t$ 关于测度 $\mathbb{Q}$ 是正态变量 $N(0,t)$。它等价于

$$\mathbb{E}_{\mathbb{Q}}[\exp(\theta\tilde{W}_t)] = \exp\left(\frac{1}{2}\theta^2 t\right)$$

(iii) $\tilde{W}_{t+s} - \tilde{W}_s$ 关于测度 $\mathbb{Q}$ 是正态变量 $N(0,t)$, 而且与 $\mathcal{F}_s$ 独立。它等价于

$$\mathbb{E}_{\mathbb{Q}}[\exp(\theta(\tilde{W}_{t+s} - \tilde{W}_s))|\mathcal{F}_s] = \exp\left(\frac{1}{2}\theta^2 t\right)$$

(ii) 和 (iii) 可由 Radon-Nikodym 导数所满足的第二个关系来验证。

3. Girsanov 定理

对一般的测度变换与Brown运动的关系, 有下面的定理。有时候也称为 Cameron-Martin-Girsanov 定理。

定理 10 设 W 关于测度 $\mathbb{P}$ 是标准Brown运动, γ_t 是可预期过程, 对 $T>0$, 满足条件 $\mathbb{E}_{\mathbb{P}}\exp\left(\frac{1}{2}\int_0^T \gamma_t^2\mathrm{d}t\right) < \infty$, 那么存在与测度 $\mathbb{P}$ 等价的测度 $\mathbb{Q}$, 使得

(i) $\dfrac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\int_0^T \gamma_t \mathrm{d}W_t - \frac{1}{2}\int_0^T \gamma_t^2 \mathrm{d}t\right)$。

(ii) $\tilde{W}_t = W_t + \int_0^t \gamma_s \mathrm{d}s$ 关于测度 $\mathbb{Q}$ 是标准Brown运动。

§6 鞅方法的应用

这一节通过一个具体的例子, 说明在一些问题中如何应用鞅方法。设 $(\Omega, \mathcal{F}, \mathbf{F}, \mathbb{P})$ 是概率空间, X 是实值随机变量, $\mathbb{E}|X| < \infty$, 那么 $M(t) = \mathbb{E}[X|\mathcal{F}_t]$ 是一个鞅。

1. 一个引理

引理 2 设 M 为连续鞅, 而且它在任意有限时间区间上是有界变差的, 那么 M 取常数, 即 $M(t) \equiv M(0)$。

证明 由 Itô 公式, 注意到有界变差性, 可得

$$M^2(t) = M^2(0) + 2\int_0^t M(s)\mathrm{d}M(s) \tag{9}$$

对 (9) 式两边取期望, 可得

$$\mathbb{E}[M^2(t)] = M^2(0)$$

因为 M 是鞅, 所以 $\mathbb{E}[M(t)] = M(0)$。从而

$$\mathrm{Var}[M(t)] = 0$$

这就证明了 M 是常数。

2. 几何 Poisson 过程

设股票在时刻 t 的价格有如下形式

$$S(t) = \mathrm{e}^{\alpha t + \beta N(t)} \tag{10}$$

其中 N 为参数为λ的 Poisson 过程。引入符号

$$V(t) = \mathbb{E}[S(t)]$$

为求 $V(t)$ 的表达式, 显然可以不必应用鞅方法。下面应用鞅方法求解, 说明有关思路。

设 $\mathcal{F}_t, t \geqslant 0$ 是由 Poisson 过程产生的滤波。选取 $T > 0$, 构造鞅 M

$$\begin{aligned} M(t) &= \mathbb{E}[S(T)|\mathcal{F}_t] \\ &= \mathrm{e}^{\alpha t + \beta N(t)}\mathbb{E}[\mathrm{e}^{\alpha(T-t)+\beta(N(T)-N(t))}|\mathcal{F}_t] \\ &= S(t)V(T-t) \end{aligned}$$

在上述推导中应用了 Poisson 过程的平稳独立增量性质。

引入符号 $V' = \dfrac{\mathrm{d}V}{\mathrm{d}t}$, 应用关于逐段可微函数的链式法则, 可得

$$\mathrm{d}M(t) = S(t)\alpha\mathrm{d}tV(T-t) + (S(t) - S(t-))V(T-t) + S(t)V'(T-t)(-\mathrm{d}t) \tag{11}$$

上式右边第二项对应于间断点。注意到

$$\begin{aligned} S(t) - S(t-) &= \mathrm{e}^{\alpha t+\beta N(t)} - \mathrm{e}^{\alpha t+\beta N(t-)} = \mathrm{e}^{\alpha t+\beta(N(t-)+\Delta N(t))} - \mathrm{e}^{\alpha t+\beta N(t-)} \\ &= S(t-)(\mathrm{e}^{\beta\Delta N(t)} - 1) \\ &= S(t-)(\mathrm{e}^{\beta} - 1)\mathrm{d}N(t) \end{aligned} \tag{12}$$

把 (12) 式代入 (11) 式, 引入鞅 M_N

$$\mathrm{d}M_N(t) = \mathrm{d}N(t) - \lambda \mathrm{d}t$$

最后得到

$$\begin{aligned}\mathrm{d}M(t) = &S(t)[\alpha V(T-t) - V'(T-t) + (\mathrm{e}^{\beta} - 1)\lambda V(T-t)]\mathrm{d}t \\ &+ S(t-)(\mathrm{e}^{\beta} - 1)\mathrm{d}M_N(t)\end{aligned} \tag{13}$$

在上式的推导中, 把 $S(t-)\mathrm{d}t$ 用 $S(t)\mathrm{d}t$ 代替, 这是因为改变函数在可列个点的取值不影响积分值。(13) 式右边第一项是两个鞅的增量之差, 从而它也是鞅的增量之差。显然, 它对应的鞅是连续的, 而且是有界变差的。由引理 2, 此鞅是常数。从而

$$\alpha V(T-t) - V'(T-t) + (\mathrm{e}^{\beta} - 1)\lambda V(T-t) = 0$$

它等价于下述微分方程

$$V'(t) = [\alpha + (\mathrm{e}^{\beta} - 1)\lambda]V(t) \tag{14}$$

由定义即得方程 (14) 的初始条件为 $V(0) = 1$。求解后, 得到

$$V(t) = \exp[(\alpha + (\mathrm{e}^{\beta} - 1)\lambda)t]$$

§7 关于半鞅的变量替换法则的一般形式

1. 关于半鞅的变量替换法则的一般形式

设 $\mathbf{X} = (X^{(1)}, \cdots, X^{(m)})$ 为 m 维半鞅, $f: \mathbb{R}^m \to \mathbb{R}$ 为 m 元二次连续可微函数。Itô 变量替换公式的一般形式为

$$\begin{aligned}\mathrm{d}f(\mathbf{X}_t) = &\sum_i \frac{\partial f}{\partial x^{(i)}}(\mathbf{X}_{t-})\mathrm{d}X_t^{(i)} + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x^{(i)}\partial x^{(j)}}(\mathbf{X}_{t-})\mathrm{d}\langle X^{(i)}, X^{(j)}\rangle_t^c \\ &+ \left\{f(\mathbf{X}_t) - f(\mathbf{X}_{t-}) - \sum_i \frac{\partial f}{\partial x^{(i)}}(\mathbf{X}_{t-})\Delta X_t^{(i)}\right\}\end{aligned} \tag{15}$$

在上述公式中, 设 $\mathbf{X}$ 的轨道是右连续的, 而且左极限存在, $\mathbf{X}_{t-}$ 表示过程在 t 点的左极限。当 $X^{(i)}$ 或 $X^{(j)}$ 是有界变差过程时, 二次变差过程 $\langle X^{(i)}, X^{(j)}\rangle \equiv 0$。对连续过程 $\mathbf{X}$, 变量替换公式仅包括右边的前两项, 进一步, 如果 $\mathbf{X}$ 的轨道是确定性的可微函数时, 那么 Itô 变量替换公式就是通常的微积分中的链式法则。

对有界变差过程, (15) 式可写为

$$\mathrm{d}f(\mathbf{X}_t) = \sum_i \frac{\partial f}{\partial x^{(i)}}(\mathbf{X}_t)\mathrm{d}(X^{(i)})_t^c + f(\mathbf{X}_t) - f(\mathbf{X}_{t-}) \tag{16}$$

特别地, 对 $f(x,y)=xy$, 由 (16) 式可得到有界变差过程的分部积分法则

$$\mathrm{d}(X_tY_t)=Y_t\mathrm{d}X_t+X_{t-}\mathrm{d}Y_t=Y_t\mathrm{d}X_t+X_t\mathrm{d}Y_t-\Delta X_t\Delta Y_t$$

为给出 Itô 变量替换公式的另一个特例, 首先简要介绍扩散过程。扩散过程是轨道连续的Markov过程, 通常这类过程可由如下随机微分方程定义

$$\mathrm{d}X_t=\kappa(t,X_t)\mathrm{d}t+\sigma(t,X_t)\mathrm{d}W_t \tag{17}$$

其中 W 是标准Brown运动, κ 称为漂移系数, σ 称为扩散系数。如果 κ 和 σ 不依赖于 X, 那么 $\mathrm{d}X_t$ 可视为正态分布变量, 均值为 $\kappa_t\mathrm{d}t$, 方差为 $\sigma_t^2\mathrm{d}t$。直观上,X 由众多独立的正态变量之和构成, 所以 X 的增量是有独立的, 而且服从正态分布,

$$X_u-X_t\sim N\left(\int_t^u\kappa_s\mathrm{d}s,\int_t^u\sigma_s^2\mathrm{d}s\right)$$

现在设 $f:\mathbb{R}\to\mathbb{R}$ 为二次连续可微函数,X 是由 (17) 式定义的扩散过程, 那么 Itô 公式变为

$$\mathrm{d}f(X_t)=f'(X_t)(\kappa(t,X_t)\mathrm{d}t+\sigma(t,X_t)\mathrm{d}W_t)+\frac{1}{2}\sigma^2(t,X_t)f''(X_t)\mathrm{d}t$$

2. 变量替换法则的一些应用

这里给出 Itô 变量替换公式的三个应用。

(i) 关于Brown运动上穿概率的一个经典结论

设 $B=\{B_t\}_{t\geqslant0}$ 为带负漂移项的Brown运动, 即 $B_t=-ct+\sigma W_t$, 其中 $c>0$, W 为标准Brown运动。给定水平 $x>0$, 考虑 B 不会达到 x 的概率, 它等价于 σW_t 不会达到直线边界 $\{c+ct\}_{t\geqslant0}$ 的概率。记此概率为 $P(x)$, 那么

$$P(x)=\mathbb{P}\left[\bigcap_{s\geqslant0}\{B_s<x\}\right]=\mathbb{P}\left[\bigcap_{s\geqslant0}\{\sigma W_s<x+cs\}\right]$$

上述问题的解为

$$P(x)=1-\exp\left(-\frac{2c}{\sigma^2}x\right) \tag{18}$$

(18) 式是关于Brown运动的一个经典结论。下面应用 Itô 公式得到 (18) 式。

记 $\mathcal{F}_t=\sigma\{B_s:s\leqslant t\}$, 构造鞅 M 如下:

$$M_t=\mathbb{P}\left[\bigcap_{s\geqslant0}\{\sigma W_s<x+cs\}|\mathcal{F}_t\right] \tag{19}$$

应用Brown运动的平稳独立增量性, 可得

$$M_t=I_tP(x+ct-\sigma W_t)$$

其中 I 为示性过程,

$$I_t = 1\left[\bigcap_{0\leqslant s\leqslant t}\{\sigma W_s < x + cs\}\right]$$

应用 Itô 变量替换公式的一般形式, 注意到示性过程 I 的连续部分为常数, 可以得到

$$\begin{aligned}\mathrm{d}M_t = \ & I_{t-}\left(P'(x+ct-\sigma W_t)(c\mathrm{d}t - \sigma \mathrm{d}W_t) + \frac{\sigma^2}{2}P''(x+ct-\sigma W_t)\mathrm{d}t\right)\\ & + I_t P(x+ct-\sigma W_t) - I_{t-}P(x+ct-\sigma W_{t-})\end{aligned}$$

由于 $I_tP(x+ct-\sigma W_t)$ 是连续的, 上式右边最后两项之差为 0。最后把含有 $\mathrm{d}W_t$ 的项移到左边, 那么左边变为两个连续鞅的差, 从而仍为鞅, 而右边变为连续的有界变差过程。因此新的鞅为常数, 由此得到

$$P'(x)c + \frac{\sigma^2}{2}P''(x) = 0$$

上述微分方程的解为

$$P(x) = a + b\exp\left(-\frac{2c}{\sigma^2}x\right)$$

应用边界条件 $P(0) = 0, P(\infty) = 1$, 可得到 (18) 式。

由 (18) 式即知Brown运动 $\{\sigma W_t\}_{t\geqslant 0}$ 上穿直线边界 $\{c+ct\}_{t\geqslant 0}$ 的概率为 $1 - P(x) = \exp\left(-\dfrac{2c}{\sigma^2}x\right)$。

对大量的独立风险的保单组合, 如果以 B 表示累积索赔与累积保费之差过程的扩散近似, 在单位时间内净收益的期望和方差分别为 c 和 σ^2, 那么 $P(x)$ 就是当初始资本为 x 时, 破产不会发生的概率。相应地, 破产概率为 $Q(x) = 1 - P(x)$。

(ii) Poisson 索赔过程下的破产概率

设索赔的发生构成一个 Poisson 过程, 强度参数为常数 λ。另外, 单个索赔额是独立同分布的, 分布函数为 G。记 $N(t, A)$ 为在时间区间 $(0, t]$ 内发生的索赔额属于 A 的范围的索赔次数。设在时刻 t 的保费率为 $c(X_t)$, 一般地它与当前盈余额 X_t 有关系。支付函数可以表示为

$$\mathrm{d}B_t = \int_y yN(\mathrm{d}t, \mathrm{d}y) - c(X_t)\mathrm{d}t \tag{20}$$

注意到以下过程是鞅:

$$\int_0^t \int_y (yN(\mathrm{d}\tau, \mathrm{d}y) - \lambda \mathrm{d}\tau G(\mathrm{d}y))$$

进一步假设利率累积因子为

$$r_t = \mathrm{e}^{\delta t + \sigma W_t} \tag{21}$$

其中 δ, σ 是常数,W 是与索赔过程独立的标准Brown运动。

首先, 设盈余过程在时刻 0 的值为 X_0, 那么它在时刻 t 的值为

$$X_t = X_0 e^{r_t} - \int_0^t e^{r_t - r_\tau} dB_\tau \tag{22}$$

对 $s > t, X_s$ 和 X_t 的关系如下

$$X_s = X_t e^{r_s - r_t} - \int_t^s e^{r_s - r_\tau} dB_\tau \tag{23}$$

现在设 $X_t = x$, 对给定的 u, 考虑在时间 $(t, u]$ 内破产不会发生的概率,

$$\mathbb{P}\left[\bigcap_{t \leqslant s \leqslant u} \{X_s > 0\} | \mathcal{F}_t\right]$$

由 (23) 式, 上述概率可表示为

$$P(t, x) = \mathbb{P}\left[\sup_{t \leqslant s \leqslant u} \int_t^s e^{-(r_\tau - r_t)} dB_\tau < x | \mathcal{F}_t\right] \tag{24}$$

为求 $P(t, x)$, 构造如下鞅

$$M_t = \mathbb{P}\left[\bigcap_{0 \leqslant s \leqslant u} \{X_s > 0\} | \mathcal{F}_t\right] \tag{25}$$

M 可以表示为

$$M_t = I_t P(t, X_t)$$

其中 I_t 表示在 $[0, t]$ 内破产不会发生的示性函数。

下面给出应用 Itô 公式的详细过程。M 又可以表示为

$$f(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}) = X^{(1)} P(X^{(2)}, e^{X^{(3)}} X^{(4)})$$

其中

$$\begin{aligned}
X_t^{(1)} &= I_t \\
X_t^{(2)} &= t \\
X_t^{(3)} &= \delta t + \sigma W_t \\
X_t^{(4)} &= X_0 - \int_0^t e^{-(\delta s - \sigma W_s^*)} \left(\int_y y N(ds, dy) - c(X_s) ds\right)
\end{aligned}$$

相关的偏导数为

$$\frac{\partial}{\partial x^{(1)}} f = P(x^{(2)}, \mathrm{e}^{x^{(3)}} x^{(4)})$$
$$\frac{\partial}{\partial x^{(2)}} f = x^{(1)} \frac{\partial}{\partial t} P(x^{(2)}, \mathrm{e}^{x^{(3)}} x^{(4)})$$
$$\frac{\partial}{\partial x^{(3)}} f = x^{(1)} \frac{\partial}{\partial x} P(x^{(2)}, \mathrm{e}^{x^{(3)}} x^{(4)}) \mathrm{e}^{x^{(3)}} x^{(4)}$$
$$\frac{\partial}{\partial x^{(4)}} f = x^{(1)} \frac{\partial}{\partial x} P(x^{(2)}, \mathrm{e}^{x^{(3)}} x^{(4)}) \mathrm{e}^{x^{(3)}}$$
$$\frac{\partial^2}{\partial x^{(3)2}} f = x^{(1)} \frac{\partial^2}{\partial x^2} P(x^{(2)}, \mathrm{e}^{x^{(3)}} x^{(4)}) \left(\mathrm{e}^{x^{(3)}} x^{(4)}\right)^2 + x^{(1)} \frac{\partial}{\partial x} P(x^{(2)}, \mathrm{e}^{x^{(3)}} x^{(4)}) \mathrm{e}^{x^{(3)}} x^{(4)}$$

应用求Brown运动上穿概率的方法, 最后可得到 $P(t,x)$ 满足如下积分微分方程:

$$\begin{aligned} &\frac{\partial}{\partial t} P(t,x) + \left(c(x) + x\left(\delta + \frac{\sigma^2}{2}\right)\right) \frac{\partial}{\partial x} P(t,x) + \frac{x^2\sigma^2}{2} \frac{\partial^2}{\partial x^2} P(t,x) \\ &+\lambda\left(P(t,x) - \int P(t,x-y) G(\mathrm{d}y)\right) = 0 \end{aligned} \tag{26}$$

需要指出, 为求上述方程的解, 需要一些边界条件, 例如需要知道在时刻 u 时 $P(t,x)$ 的取值。如果令 $u \to \infty$, 那么就得到在无限时间内破产不会发生的概率, 它也是 $[0,\infty)$ 内破产不会发生的概率, 记为 $P(x)$,$P(x)$ 满足如下方程

$$\begin{aligned} &\left(c(x) + x\left(\delta + \frac{\sigma^2}{2}\right)\right) P'(x) + \frac{x^2\sigma^2}{2} P''(x) \\ &+\lambda\left(P(x) - \int P(x-y) G(\mathrm{d}y)\right) = 0 \end{aligned} \tag{27}$$

进一步, 如果利率是确定性的, 即 $\sigma = 0$, 而且保费率是常数 c, 那么 (27) 式变为

$$(c + x\delta) P'(x) + \lambda\left(P(x) - \int P(x-y) G(\mathrm{d}y)\right) = 0 \tag{28}$$

最后, 如果不考虑利率, 即 $r_t \equiv 0$, 那么 $P(x)$ 满足的方程可进一步简化如下

$$cP'(x) + \lambda\left(P(x) - \int P(x-y) G(\mathrm{d}y)\right) = 0 \tag{29}$$

(29) 式就是经典风险理论的结论。

(iii) 一个永久年金的概率分布

考虑如下随机变量

$$X = \int_0^\infty \mathrm{e}^{-\delta\tau - \sigma W_\tau} \mathrm{d}\tau \tag{30}$$

X 可视为利率累积因子为 $r_t = \mathrm{e}^{\delta t+\sigma W_t}$ 下的连续永久年金的现值。为求 X 的概率分布, 注意到

$$\begin{aligned}\mathbb{P}[X < x] &= \mathbb{P}\left[\bigcap_{t\geqslant 0}\left\{\int_0^t \mathrm{e}^{-\delta\tau-\sigma W_\tau}\mathrm{d}\tau < x\right\}\right]\\ &= \mathbb{P}\left[\bigcap_{t\geqslant 0}\left\{x-\int_0^t \mathrm{e}^{-\delta\tau-\sigma W_\tau}\mathrm{d}\tau > 0\right\}\right]\\ &= P(x)\end{aligned} \tag{31}$$

因此 (31) 式表明 $\mathbb{P}[X < x]$ 可视为特殊情形下的生存概率 $P(x)$。此时不存在索赔, 保费率 $c = -1$, 初始值为 x, 存在随机利率时, 从而 (27) 式变为

$$\left(-1+x\left(\delta+\frac{\sigma^2}{2}\right)\right)P'(x)+\frac{x^2\sigma^2}{2}P''(x)=0 \tag{32}$$

由 (32) 式可得

$$P'(x)=C_1\frac{1}{x^{(2\delta/\sigma^2)+1}}\exp\left(-\frac{2}{\sigma^2 x}\right) \tag{33}$$

其中 C_1 是某系数。

由 (33) 式可知, $Y = X^{-1}$ 的概率密度函数为

$$C_2 y^{(2\delta/\sigma^2)-1}\exp\left(-\frac{y}{\sigma^2/2}\right) \tag{34}$$

其中 C_2 是某系数。由 (34) 式, X^{-1} 服从 Gamma 分布, 形状参数为 $2\delta/\sigma^2$, 标度参数为 $\sigma^2/2$。由此即得当 $k < 2\delta/\sigma^2$ 时

$$\mathbb{E}[X^k]=\left(\prod_{j=1}^{k}[\delta-(j\sigma^2)/2]\right)^{-1}$$

特别地, 当 $\sigma^2 < 2\delta$ 时, $\mathbb{E}[X]=(\delta-\sigma^2/2)^{-1}$。注意到 $\mathbb{E}[X]$ 关于 σ^2 是单增的, $\mathbb{E}[X] > \delta^{-1}$。回顾一下, 确定性利率力 δ 的永久年金的现值就是 δ^{-1}。因此, 如果利率力是随机的, 那么永久年金的现值变量的期望值应变大。

第七章　寿险中的随机性

§1　寿险数学的基本概念

1. 引言

本节从随机过程的观点阐述寿险数学。寿险与相关的保险依赖于生命事件 (如疾病、死亡等), 它们按一定次序构成了个体的生命历史, 这正是应用随机过程阐述寿险数学的内在原因。生命历史可视为一个适当的随机过程的样本路径。最简单的生命事件是死亡, 此时即使不应用随机过程, 也可以建立死亡模型 (例如, 把剩余寿命视为随机变量), 而其他的生命事件就不再这么简单, 因此在建立模型时, 需要考虑整个生命历史。很自然地, 随机过程的理论和方法就构成了一个出发点。

为清楚地说明思路, 首先引入一个简单的例子。如图 1 所示, 在此例中, 两个状态模型 (或单重损失模型) 以很直观的方式给出, 它是寿险精算中最基本的模型。在这个例子中应用随机过程, 似乎是用一种不太熟悉的语言描述熟知的结果。对于这个事件本身, 应用随机过程并没有得到新的结果, 但需要强调的是, 当建立复杂的模型时, 随机过程的优势才显示出来。



图 1　两个状态模型

(i) 在简单过程中所发展出来的方法, 对于更为复杂的过程同样适用, 如运用于疾病或长期治疗的模型中。

(ii) 这些有用的工具也是金融数学中的工具。特别地, 随机积分和条件期望是关键性的思路与工具。

金融数学和寿险数学主要区别在于：前者是以连续路径的过程为基础, 而后者是以有跳跃的过程为基础。寿险数学中的基本研究对象是计数过程, 它是一类特殊的随机过程。

2. 计数过程

图 1 表示两个状态的Markov过程, 转移强度 μ_x 表示年龄 x 时的死亡力。为了方便, 设生存状态为 0, 死亡状态为 1, 那么这个过程的一条样本路径可由图 2 给出。在图 2 中, 某人死于 46 岁。样本路径是时间 t 的函数, 记为 $N_{01}(t)$, 由 $N_{01}(t)$ 可知在时刻 t 时死亡是否已经发生。从另一方面看, $N_{01}(t)$ 表示到 t 时已经发生的事件次数。

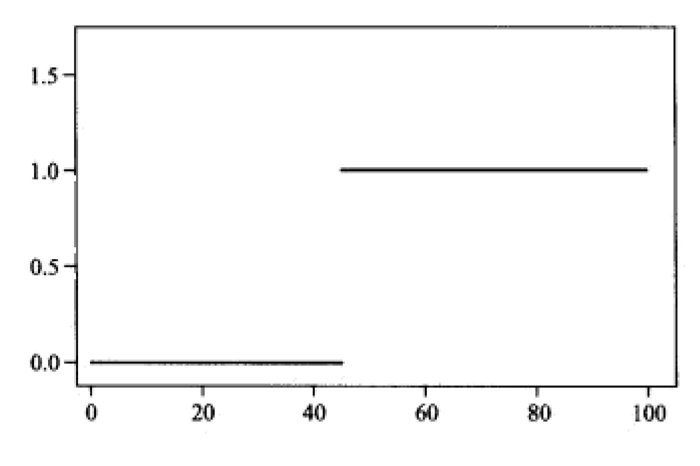


图 2　死亡在 46 岁发生的计数过程的样本路径

进一步, 考虑样本路径 $N_{01}(t)$ 的增量。在上例中, 如果在时刻 t 时过程没有出现跳跃, 那么增量为 0, 记为 $\mathrm{d}N_{01}(t)=0$；如果在时刻 t 时过程出现跳跃, 那么增量为 1, 记为 $\mathrm{d}N_{01}(t)=1$。有时, 也用 $\Delta N_{01}(t)$ 表示 $\mathrm{d}N_{01}(t)$。

对于计数过程来说, 增量 $\mathrm{d}N_{01}(t)$ 类似于可微的样本路径的一阶导数。可微的样本路径, 可通过对导数取积分构造出来。相应地, 计数过程的样本路径, 可通过对增量取积分构造出来。这就引出了随机积分的概念。

3. 随机积分

首先, 考虑跳跃仅发生在整数时刻的一般的离散时间计数过程。因此, 在所有非整数时刻, $\mathrm{d}N_{01}(t)=0$, 而在某些整数时刻, $\mathrm{d}N_{01}(t)=1$。能否由增量 $\mathrm{d}N_{01}(t)$ 构造 $N_{01}(t)$? 更具体地说, 能否得到 $N_{01}(T)$(T 不必是整数)? 设 $J(T)$ 表示到 T 时的

所有可能的发生跳跃的时刻点，那么

$$N_{01}(T)=\sum_{t\in J(T)}\mathrm{d}N_{01}(t) \tag{1}$$

其次，假设 $N_{01}(t)$ 仍是离散时间过程，但在更多的时刻点可能发生跳跃，如以月为单位，在每个月末可能发生跳跃，$J(T)$ 仍表示到 T 时的所有可能的发生跳跃的时刻点，此时等式 (1) 仍然成立。

进一步，如果时间单位越来越短，那么取极限后，就得到以下结论。

(i) 计数过程是连续时间的过程。

(ii) 可能发生跳跃的时刻点集合是区间 $(0,T]$。

(iii) 求和形式变为积分形式，

$$N_{01}(T)=\int_{t\in J(T)}\mathrm{d}N_{01}(t)=\int_0^T\mathrm{d}N_{01}(t) \tag{2}$$

(2) 式的积分就是一个随机积分。当视为 T 的函数时，(2) 式就构成一个随机过程。

4. 保险与年金

考虑在死亡时刻支付 1 个单位金额的终身寿险。对 (x)，以 X 表示此终身寿险的现值。如果以 T_x 表示 (x) 的剩余寿命随机变量，那么在寿险精算中，$X=v^{T_x}=\mathrm{e}^{-\delta T_x}$。

如果应用随机积分来表示 X，注意到在时刻 t 时支付额为 1 个单位的现值是 v^t，进一步，如果在时刻 t 时没有死亡发生，那么计数过程的增量为 $\mathrm{d}N_{01}(t)=0$，现值就是 $v^t\mathrm{d}N_{01}(t)=0$。如果在时刻 t 时死亡发生，那么计数过程的增量为 $\mathrm{d}N_{01}(t)=1$，现值就是 $v^t\mathrm{d}N_{01}(t)=v^t$。求积分后可以得到

$$X=\int_0^\infty v^t\mathrm{d}N_{01}(t) \tag{3}$$

对年金同样有随机积分表示。考虑连续支付的生命年金，设支付率为 1，以 Y 表示年金现值。构造如下随机过程：如果在时刻 t 时没有死亡发生，那么 $I_0(t)=1$，否则 $I_0(t)=0$。称 $\{I_0(t):t\geqslant 0\}$ 为示性过程。由此可得

$$Y=\int_0^\infty v^t I_0(t)\mathrm{d}t \tag{4}$$

对给定的样本路径来说，(4) 式是通常的积分形式。由于样本路径是随机的，所以 Y 是随机变量。如果以 $X(T)$ 和 $Y(T)$ 分别表示到 T 时的支付的现值，那么以上两式可修正为

$$X(T)=\int_0^T v^t\mathrm{d}N_{01}(t),\quad Y(T)=\int_0^T v^t I_0(t)\mathrm{d}t \tag{5}$$

5. 寿险数学基础

由前面的例子, 现在可以从计数过程的观点来介绍寿险数学的基本内容。首先引入一般的支付函数,

(i) 如果 $N(t)=0$, 即生存状态, 那么年金的支付率为 $a_0(t)$。

(ii) 如果在时刻 t 时 N 由 0 跳跃到 1, 即死亡发生, 那么保险的保额为 $A_{01}(t)$。

注意到, 相对于保险人而言, 保费可视为取负值的年金, 而且以上支付函数可以推广到多状态模型。同样地, 支付函数也可以包含离散年金或纯生保险的保额。为使讨论简便, 这里对离散支付年金不作讨论。

$a_0(t)$ 和 $A_{01}(t)$ 都是时间的函数, 但不必是随机过程, 它们对应于不同的事件的支付, 但并不表示事件本身, 在通常的保险中, 它们是年龄的函数。支付率可以表示为如下 $\mathrm{d}L(t)$ 的形式

$$\mathrm{d}L(t)=A_{01}(t)\mathrm{d}N_{01}(t)+a_0(t)I_0(t)\mathrm{d}t \tag{6}$$

上式给出了在 $(t,t+\mathrm{d}t)$ 内的支付率。假设时刻 T 之后没有支付量 (T 可为 ∞), 那么总支付量为

$$L=\int_0^T\mathrm{d}L(t)=\int_0^T A_{01}(t)\mathrm{d}N_{01}(t)+\int_0^T a_0(t)I_0(t)\mathrm{d}t \tag{7}$$

如以 $V(0)$ 表示总支付在时刻 0 的价值, 那么

$$V(0)=\int_0^T v^t\mathrm{d}L(t)=\int_0^T v^tA_{01}(t)\mathrm{d}N_{01}(t)+\int_0^T v^ta_0(t)I_0(t)\mathrm{d}t \tag{8}$$

把 (8) 式与 (5) 式比较, 可知 (8) 式考虑到更一般的支付, 它表示取决于特定的生命历史的支付, 即支付对应于样本路径 $N_{01}(t)$。作为 T 的函数, (8) 式定义了一个随机过程。

如以 $V(s)$ 表示在任意时刻s的价值, 那么

$$V(s)=\frac{1}{v^s}\int_0^T v^t\mathrm{d}L(t)=\frac{1}{v^s}\int_0^T v^tA_{01}(t)\mathrm{d}N_{01}(t)+\frac{1}{v^s}\int_0^T v^ta_0(t)I_0(t)\mathrm{d}t \tag{9}$$

需要指出, 在以上讨论中, 假设利率是常数。实际上同样可以假设利率是时间的函数, 甚至可以是随机过程。

6. 现值变量的期望

到此, 只定义了样本空间的元素, 即样本路径 $N_{01}(t)$, 以及一些相关的函数, 如 L 和 $V(s)$, 但还没有引入 σ– 代数、滤波、概率测度, 从而也就无法进行概率运算, 如计算期望等。下面将讨论这些问题。

(i) 此处的滤波是由过程 $N_{01}(t)$ 产生的"自然的"滤波, 对它的描述如下：在时刻 t 时, 对所有的 $s \leqslant t$, 考虑到 $N_{01}(s)$ 的所有可能值, 进而这些可能值构成的所有事件, 这些信息就构成了 σ- 代数 $\mathcal{F}_t$。

(ii) 总体 σ- 代数 $\mathcal{F}$ 是所有 $\mathcal{F}_t$ 的并。

(iii) 概率测度对应于生命表。根据不同的目的, 精算师会选择不同的生命表, 即样本空间和滤波与概率测度的选择是独立的。

所有的计算都依赖于概率测度 (生命表) 的选择, 这可由现值变量的期望来说明。假设精算师选择概率测度 $\mathbb{P}$, 它等价于生命表概率 ${}_tp_x$, 考虑关于 (x) 的终身寿险, $\mathbb{E}_{\mathbb{P}}[X]$ 为

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[\int_0^{\infty} v^t \mathrm{d}N_{01}(t)\right] &= \int_0^{\infty} v^t \mathbb{E}_{\mathbb{P}}[\mathrm{d}N_{01}(t)] = \int_0^{\infty} v^t \mathbb{P}[\mathrm{d}N_{01}(t)=1] \\ &= \int_0^{\infty} v^t\ {}_tp_x \mu_{x+t} \mathrm{d}t\end{aligned} \tag{10}$$

(10) 式是寿险精算中的常识。如果精算师选择另一个概率测度 $\mathbb{P}^*$, 它等价于另一生命表概率 ${}_tp_x^*$, 那么可以得到另一个不同的期望

$$\mathbb{E}_{\mathbb{P}^*}[X] = \int_0^{\infty} v^t\ {}_tp_x^* \mu_{x+t}^* \mathrm{d}t \tag{11}$$

类似地, 年金变量的期望为

$$\mathbb{E}_{\mathbb{P}}[Y] = \int_0^{\infty} v^t\ {}_tp_x \mathrm{d}t \tag{12}$$

7. 关于计数过程的其他例子

图 3 给出了一个常见的生病—死亡模型。以 $S(t)$ 表示时刻 t时的状态, 这里有三个状态：健康、生病、死亡, 分别记为 0, 1, 2。图 4 给出了 $S(t)$ 的一个样本路径：某人在 40 岁时生病, 到 42 岁时恢复健康。其后在 49 岁时再次生病, 直到在 60 岁时死亡。

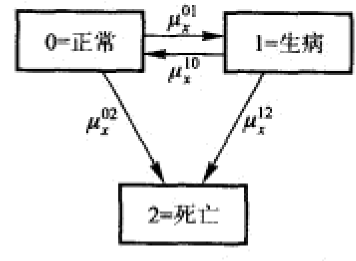


图 3 生病 — 死亡模型

对上述生病—死亡模型, 可以定义四个计数过程, 每个计数过程都对应于状态的转变次数, 例如 $N_{01}(t)$ 表示从状态 0 转变为状态 1 的次数, 类似地可解释 $N_{02}(t), N_{10}(t), N_{12}(t)$ 的意义。

如果把以上作为一个整体来考虑, 那么可得到含有四个部分的计数过程。另外, 可以定义示性随机过程 $I_j(t)$, 依赖于状态的年金支付函数 $a_j(t)$, 以及依赖于状态转移的保额函数 $A_{jk}(t)$ 等。只需改变一些符号, 关于二状态模型的寿险数学就可以很容易地推广到多状态模型中。

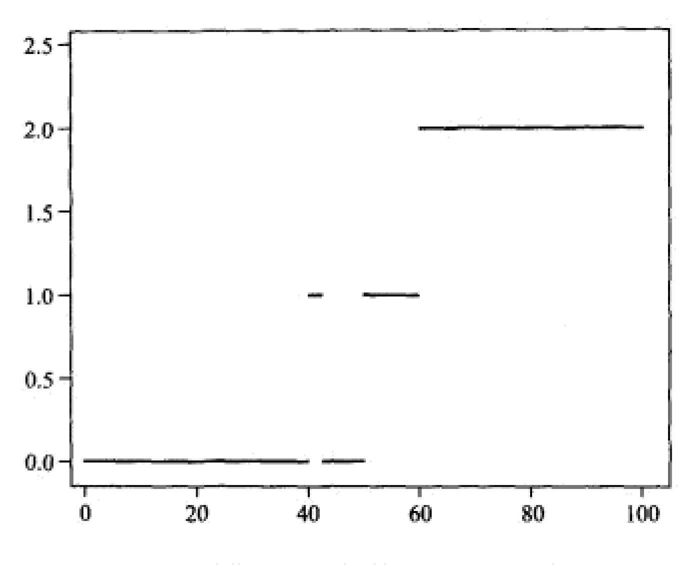


图 4　生病 — 死亡过程的一条样本路径

8. 鞅

至此还没有涉及与计数过程有关的鞅。下面的鞅非常简单, 在应用中起着很重要的作用。在二状态模型中的鞅为

$$M_{01}(t)=N_{01}(t)-\int_0^t I_0(s)\mu_{x+s}\mathrm{d}s \tag{13}$$

称 $M_{01}(t)$ 为补偿计数过程, 而等式右边的积分项称为 $N_{01}(t)$ 的补偿因子。为说明 $M_{01}(t)$ 是鞅, 注意到 $M_{01}(t)$ 的增量有以下性质

$$\mathrm{d}M_{01}(t)=\mathrm{d}N_{01}(t)-I_0(t)\mu_{x+t}\mathrm{d}t \tag{14}$$

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[\mathrm{d}M_{01}(t)]&=\mathbb{E}_{\mathbb{P}}[\mathrm{d}N_{01}(t)]-\mathbb{E}_{\mathbb{P}}[I_0(t)\mu_{x+t}\mathrm{d}t]\\&={}_t\,p_x\mu_{x+t}-{}_t\,p_x\mu_{x+t}=0\end{aligned} \tag{15}$$

在上述期望中涉及概率测度 $\mathbb{P}$。如果改变概率测度, 如选择对应于 ${}_tp_x^*$ 的概率测度 $\mathbb{P}^*$, 那么可得到另外一个鞅

$$M_{01}^*(t)=N_{01}(t)-\int_0^t I_0(s)\mu_{x+s}^*\mathrm{d}s \tag{16}$$

此外, 给定死亡力 μ_t^*, 也可以得到概率测度 $\mathbb{P}^*$, 使得 $M_{01}^*(t)$ 是 $\mathbb{P}^*$ 鞅。实际上, $\mathbb{P}^*$ 对应于 ${}_tp_x^*=\exp\left(-\int_0^t\mu_{x+s}^*\mathrm{d}s\right)$。

对于更为复杂的模型, 可以得到一系列鞅, 每个鞅对应于从状态的转变, 鞅的形式为

$$M_{jk}(t)=N_{jk}(t)-\int_0^t I_j(s)\mu_{x+s}^{jk}\mathrm{d}s \tag{17}$$

§2 逐段可微函数与积分

1. 逐段可微函数

为了描述某现象的变化, 可考虑定义于时间区间 $[0,\infty)$ 上的函数 $X=(X_t)_{t\geqslant 0}$, X_t 为在时刻t时的状态或取值。在以后的讨论中, 设 X 的取值为实数。

首先设函数 X 是有界变差的, 即它可以表示为两个非减函数之差。此时, 由有界单调序列的收敛性, 可知对所有的 t, 左极限和右极限都是存在的, 分别记为 $X_{t-}=\lim\limits_{s\uparrow t}X_s$ 和 $X_{t+}=\lim\limits_{s\downarrow t}X_s$。另外, 可以证明 X 的不连续点集 $\mathcal{D}(X)$ 至多有可列个。

特别地, 考虑有界变差函数类中的子类, 此时要求函数是右连续的而且左极限存在, 简记为 RCLL。概率分布函数属于这个子类, 另外金融保险中的支付量函数也属于这个子类。如果 X 属于这个子类, 而且 $\Delta X_t=X_t-X_{t-}$ 不为零, 那么称 ΔX 为 X 在时刻 t 的跳跃度。

以下考虑具有如下形式的 X

$$X_t=X_0+\int_0^t x_\tau \mathrm{d}\tau+\sum_{0<\tau\leqslant t}(X_\tau-X_{\tau-}) \tag{18}$$

(18) 式右边的积分项表示对连续增量求和, 求和项表示对不连续点的跳跃度求和。

进一步, 假设 X 是逐段可微的。显然, X 是逐段可微的等价于 X 和 x 都是逐段连续的。此时, 在每一点 $t\notin\mathcal{D}=\mathcal{D}(x)\cup\mathcal{D}(X)$, $\dfrac{\mathrm{d}X_t}{\mathrm{d}t}=x_t$, X 在时刻 t 的连续增长率为 x_t。

(18) 式可写为如下简便的微分形式

$$\mathrm{d}X_t=x_t\mathrm{d}t+X_t-X_{t-} \tag{19}$$

2. 关于函数的积分

设 X 和 Y 都是逐段可微函数, 而且 X 可写为 (19) 式。在时间区间 $(s,t]$ 上 Y 关于 X 的积分定义为

$$\int_s^t Y_\tau \mathrm{d}X_\tau=\int_s^t Y_\tau x_\tau \mathrm{d}\tau+\sum_{s<\tau\leqslant t}Y_\tau(X_\tau-X_{\tau-}) \tag{20}$$

(20) 式定义的积分可视为t的函数, 它是逐段可微的, 它在时刻 t 的连续增量为

$Y_t x_t \mathrm{d}t$, 跳跃度为 $Y_t(X_t - X_{t-})$。特别地, 取 $Y \equiv 1$, (20) 式变为

$$\begin{aligned}\int_s^t \mathrm{d}X_\tau &= \int_s^t x_\tau \mathrm{d}\tau + \sum_{s<\tau\leqslant t}(X_\tau - X_{\tau-}) \\ &= \int_0^t x_\tau \mathrm{d}\tau + \sum_{0<\tau\leqslant t}(X_\tau - X_{\tau-}) - \left(\int_0^s x_\tau \mathrm{d}\tau + \sum_{0<\tau\leqslant s}(X_\tau - X_{\tau-})\right) \\ &= X_t - X_s\end{aligned}$$

上式最后一式用到 (18) 式, 从而

$$X_t = X_s + \int_s^t \mathrm{d}X_\tau \tag{21}$$

下面顺便给出另外两种积分形式

$$\int_{s-}^t Y_\tau \mathrm{d}X_\tau = \lim_{r\uparrow s}\int_r^t Y_\tau \mathrm{d}X_\tau = \int_{[s,t]} Y_\tau \mathrm{d}X_\tau$$
$$\int_s^{t-} Y_\tau \mathrm{d}X_\tau = \lim_{r\uparrow t}\int_s^r Y_\tau \mathrm{d}X_\tau = \int_{(s,t)} Y_\tau \mathrm{d}X_\tau$$

3. 链式法则

设 $\mathbf{X}_t = (X_t^1, \cdots, X_t^m)$ 是 m 元函数, 每个分量都是逐段可微的, $\mathrm{d}X_t^i = x_t^i \mathrm{d}t + (X_t^i - X_{t-}^i)$。设 $f: \mathbb{R}^m \to \mathbb{R}$ 是实值函数, 而且有连续偏导数, 记为 $D_i f$。构造复合函数 $f(\mathbf{X}_t) = f(X_t^1, \cdots, X_t^m)$。如果在关于时间的某开区间内, x^i 是连续的, 而且 X^i 没有跳跃点, 那么 $f(\mathbf{X}_t)$ 满足通常的链式法则; 而在例外点上, $f(\mathbf{X}_t)$ 有跳跃, 跳跃度为 $f(\mathbf{X}_t) - f(\mathbf{X}_{t-})$。从而下述链式法则 (或称为 Itô 公式) 成立,

$$\mathrm{d}f(\mathbf{X}_t) = \sum_{i=1}^m D_i f(\mathbf{X}_t) x_t^i \mathrm{d}t + f(\mathbf{X}_t) - f(\mathbf{X}_{t-}) \tag{22}$$

(22) 式的积分形式为

$$f(\mathbf{X}_t) = f(\mathbf{X}_s) + \int_s^t \sum_{i=1}^m D_i f(\mathbf{X}_\tau) x_\tau^i \mathrm{d}t + \sum_{s<\tau\leqslant t}\{f(\mathbf{X}_\tau) - f(\mathbf{X}_{\tau-})\} \tag{23}$$

显然, $f(\mathbf{X}_t)$ 是逐段可微的。

一个常见的特例由下式给出

$$\begin{aligned}\mathrm{d}(X_t Y_t) &= X_t y_t \mathrm{d}t + Y_t x_t \mathrm{d}t + X_t Y_t - X_{t-} Y_{t-} \\ &= X_{t-}\mathrm{d}Y_t + Y_{t-}\mathrm{d}X_t + (X_t - X_{t-})(Y_t - Y_{t-}) \\ &= X_{t-}\mathrm{d}Y_t + Y_t \mathrm{d}X_t\end{aligned} \tag{24}$$

当 X 和 Y 没有共同的跳跃点时, (24) 式变为通常的形式

$$\mathrm{d}(X_tY_t) = X_t\mathrm{d}Y_t + Y_t\mathrm{d}X_t \tag{25}$$

(24) 式也可写为如下分部积分形式

$$\int_s^t Y_\tau \mathrm{d}X_\tau = Y_tX_t - Y_sX_s - \int_s^t X_{\tau-}\mathrm{d}Y_\tau \tag{26}$$

4. 一些特殊情形

现在考虑 (26) 式的三种特殊情形。不失一般性, 在 (26) 式中设 $s=0$, 可得

$$Y_tX_t = Y_0X_0 + \int_0^t Y_\tau \mathrm{d}X_\tau + \int_0^t X_{\tau-}\mathrm{d}Y_\tau \tag{27}$$

(27) 式表明 XY 在时刻 t 的值等于 XY 在时刻 0 的值与区间 $(0,t]$ 内的增量之和。下面分三种情况考察 (27) 式。

(i) 设 X 和 Y 都是离散的。为简化符号, 设 $X_t = \sum_{j=0}^{[t]} x_j, Y_t = \sum_{j=0}^{[t]} y_j$。那么

$$\begin{aligned}
X_tY_t &= \sum_{j=0}^{[t]} x_j \sum_{j=0}^{[t]} y_j = \sum_{i,j} x_iy_j = x_0y_0 + \sum_{i\geqslant j, i\geqslant 1} x_iy_j + \sum_{i<j} x_iy_j \\
&= x_0y_0 + \sum_{i=1}^{[t]}\sum_{j=0}^{i} y_jx_i + \sum_{j=1}^{[t]}\sum_{i=0}^{j-1} x_iy_j \\
&= x_0y_0 + \sum_{i=1}^{[t]} Y_ix_i + \sum_{j=1}^{[t]} X_{j-1}y_j \\
&= x_0y_0 + \int_0^t Y_\tau\mathrm{d}X_\tau + \int_0^t X_{\tau-}\mathrm{d}Y_\tau
\end{aligned}$$

上式就是 (27) 式。在上面最后表达式中, 求和项表示成广义形式的积分。

(ii) 如果 X 和 Y 都是连续的,

$$X_t = X_0 + \int_0^t x_\tau \mathrm{d}\tau,\ Y_t = Y_0 + \int_0^t y_\tau \mathrm{d}\tau$$

设 $X_0=Y_0=0$, 那么

$$\begin{aligned}X_tY_t&=\int_0^t x_\sigma\mathrm{d}\sigma\int_0^t y_\tau\mathrm{d}\tau\\&=\iint_{0<\tau\leqslant\sigma\leqslant t}y_\tau\mathrm{d}\tau x_\sigma\mathrm{d}\sigma+\iint_{0<\sigma<\tau\leqslant t}x_\sigma\mathrm{d}\sigma y_\tau\mathrm{d}\tau\\&=\int_0^t\int_0^\sigma y_\tau\mathrm{d}\tau x_\sigma\mathrm{d}\sigma+\int_0^t\int_0^{\tau-}x_\sigma\mathrm{d}\sigma y_\tau\mathrm{d}\tau\\&=\int_0^t Y_\sigma x_\sigma\mathrm{d}\sigma+\int_0^t X_{\tau-}y_\tau\mathrm{d}\tau\\&=\int_0^t Y_\tau\mathrm{d}X_\tau+\int_0^t X_\tau\mathrm{d}Y_\tau\end{aligned}$$

上式就是 (27) 式。在上面最后表达式中, 因为改变被积函数在可列个点的数值, 积分值不变, 因此可把 $X_{\tau-}$ 换为 X_τ。

(iii) 设 X 和 Y 有一个是连续的, 另一个是离散的。如 $X_t=\sum\limits_{j=0}^{[t]}x_j, Y_t=Y_0+\int_0^t y_\tau\mathrm{d}\tau$。引入

$$\hat{y}_0=Y_0,\ \hat{y}_j=\int_{j-1}^j y_\tau\mathrm{d}\tau,\ j=1,\cdots,[t]$$

那么 $X_t=X_{[t]}$, 而且

$$\begin{aligned}X_tY_t&=X_{[t]}Y_{[t]}+X_{[t]}(Y_t-Y_{[t]})\\&=\sum_{i=0}^{[t]}x_i\sum_{j=0}^{[t]}\hat{y}_j+X_{[t]}\int_{[t]}^t y_\tau\mathrm{d}\tau\end{aligned}\tag{28}$$

应用情形 (i) 中的结论, (28) 式右边第一项为

$$\begin{aligned}&X_0Y_0+\sum_{i=1}^{[t]}\sum_{j=0}^{i}\hat{y}_jx_i+\sum_{j=1}^{[t]}\sum_{i=0}^{j-1}x_i\hat{y}_j\\&=X_0Y_0+\sum_{i=1}^{[t]}Y_ix_i+\sum_{j=1}^{[t]}X_{j-1}\int_{j-1}^j y_\tau\mathrm{d}\tau\\&=X_0Y_0+\int_0^t Y_\tau\mathrm{d}X_\tau+\int_0^{[t]}X_{\tau-}\mathrm{d}Y_\tau\end{aligned}$$

而 (28) 式右边第二项为 $\int_{[t]}^t X_{\tau-}\mathrm{d}Y_\tau$。因此, 在这种情形下, (27) 式也成立。

5. 计数过程

设 $t_1 < t_2 < \cdots$ 为 $(0,\infty)$ 内的一个序列, 而且 $\lim\limits_{j\to\infty} t_j = \infty$。$t_j$ 可解释为某事件第 j 次发生的时间。在时间区间 $[0,t]$ 内事件发生的次数记为 $N_t = \#\{j; t_j \leqslant t\} = \sum\limits_j 1_{[t_j,\infty)}(t)$。如记 $t_0 = 0$, 那么当 $t_j \leqslant t < t_{j+1}$ 时, $N_t = j$。这样定义的函数 $N = (N_t)_{t\geqslant 0}$ 称为计数函数, 它是一个特殊的逐段可微函数。

在 (23) 式中, 取 X 为计数函数 N, 而且 $f: \mathbb{R} \to \mathbb{R}$, 可得

$$\begin{aligned} f(N_t) &= f(N_s) + \sum_{s<\tau\leqslant t} \{f(N_\tau) - f(N_{\tau-})\} \\ &= f(N_s) + \sum_{s<\tau\leqslant t} \{f(N_{\tau-}+1) - f(N_{\tau-})\}(N_\tau - N_{\tau-}) \\ &= f(N_s) + \int_s^t \{f(N_{\tau-}+1) - f(N_{\tau-})\}\mathrm{d}N_\tau \end{aligned} \tag{29}$$

注意到 (29) 式和下式是类似的

$$f(j) = f(0) + \sum_{i=1}^{j} \{f(i) - f(i-1)\}$$

§3 支付量函数

1. 支付量函数

在金融与保险中, 为描述在一段时间内有效的合同所导致的现金流变化, 可引入支付量函数。设在时刻 0 时合同生效, 在时刻 T 中止。记 A_t 为在区间 $[0,t]$ 内的总支付量, 支付量函数 $(A_t)_{t\geqslant 0}$ 可表示为两个非减函数之差, 它们分别表示收入与支出。因此, $(A_t)_{t\geqslant 0}$ 是有界变差函数, 进一步假设它是右连续的, 而且左极限存在。为了在数学上处理方便, 可设 $(A_t)_{t\geqslant 0}$ 是逐段可微的, 即它有如下表示

$$A_t = A_0 + \int_0^t a_\tau \mathrm{d}\tau + \sum_{0<\tau\leqslant t} \Delta A_\tau \tag{30}$$

其中 $\Delta A_\tau = A_\tau - A_{\tau-}$。(1) 式右边积分项表示连续支付, 而求和项表示各次总付的支付。(1) 式的微分形式为

$$\mathrm{d}A_t = a_t \mathrm{d}t + \Delta A_t \tag{31}$$

在保险中, 为强调保险人对被保险人的义务, 支付量函数表示为支出与收入的差, 为使符号简便, 引入 $B = -A$。

2. 利率

支付量按照给定的利率累积。在时刻 s 时的一个单位的存款账户, 它在将来时刻 t 时的累积值记为 $S(s,t)$。对 $s<t<u$, 考虑到实际问题的意义, 下面的关系成立

$$S(s,u)=S(s,t)S(t,u) \tag{32}$$

在 (32) 式中取 $s=t=u$, 即得 $S(t,t)=1$。对固定的 t 和 u, 由 (32) 式可得, 对所有的 s, 下面的关系成立,

$$S(t,u)=\frac{S(s,u)}{S(s,t)}$$

特别地, 取 $s=0$, 而且记 $S_t=S(0,t)$, 可得

$$S(t,u)=\frac{S_u}{S_t} \tag{33}$$

称 S_t 为累积函数。相应地, S_t^{-1} 是时刻 t 时的一个单位在时刻 0 时的价值, 称为贴现函数。

在以下讨论中, 设 S_t 有如下形式

$$S_t=\mathrm{e}^{\int_0^t r}, S_t^{-1}=\mathrm{e}^{-\int_0^t r} \tag{34}$$

其中 r_t 是取正值的逐段连续函数, 称 r_t 为在时刻 t 的利率强度或瞬时利率。为表示简便, 记 $\int r=\int r_\tau \mathrm{d}\tau$。

由 (34) 式可得下面的微分形式,

$$\mathrm{d}S_t=S_t r_t \mathrm{d}t \tag{35}$$

$$\mathrm{d}S_t^{-1}=-S_t^{-1} r_t \mathrm{d}t \tag{36}$$

(35) 式的直观解释是: 在很小的区间内产生的利息, 与当时的累积值 S_t, 瞬时利率 r_t, 以及区间长度 $\mathrm{d}t$ 成比例。对 (35) 式两边取积分, 可得

$$S_t=S_s+\int_s^t S_\tau r_\tau \mathrm{d}\tau,\ s\leqslant t \tag{37}$$

对 (36) 式两边取积分, 可得 $S_u^{-1}=S_t^{-1}-\int_t^u S_\tau^{-1} r_\tau \mathrm{d}\tau$, 或表示为

$$S_t^{-1}=S_u^{-1}+\int_t^u S_\tau^{-1} r_\tau \mathrm{d}\tau,\ t\leqslant u \tag{38}$$

当瞬时利率为常数 r 时, $S_t=\mathrm{e}^{rt}, S_t^{-1}=\mathrm{e}^{-rt}$。引入年利率 i 和年贴现因子 v 如下

$$i=\mathrm{e}^r-1 \tag{39}$$

$$v = e^{-r} = \frac{1}{1+i} \tag{40}$$

此时,

$$S_t = (1+i)^t, S_t^{-1} = v^t \tag{41}$$

3. 支付量的价值评估与准备金的概念

由 (33) 式, 在时刻τ附近的小区间内的支付量 $\mathrm{d}A_\tau$ 在时刻 t 的价值为 $e^{\int_0^t r}e^{-\int_0^\tau r}\mathrm{d}A_\tau$, 关于变量 τ 取积分, 可得到所有支付量在时刻 t 的价值为

$$e^{\int_0^t r}\int_{0-}^T e^{-\int_0^\tau r}\mathrm{d}A_\tau = e^{\int_0^t r}\left(\int_{0-}^t e^{-\int_0^\tau r}\mathrm{d}A_\tau + \int_t^T e^{-\int_0^\tau r}\mathrm{d}A_\tau\right) = U_t - V_t$$

其中

$$U_t = e^{\int_0^t r}\int_{0-}^t e^{-\int_0^\tau r}\mathrm{d}A_\tau = \int_{0-}^t e^{\int_\tau^t r}\mathrm{d}A_\tau \tag{42}$$

$$V_t = e^{\int_0^t r}\int_t^T e^{-\int_0^\tau r}\mathrm{d}A_\tau = \int_t^T e^{-\int_t^\tau r}\mathrm{d}B_\tau \tag{43}$$

因此 U_t 是时刻 t 之前的收入与支出的差在时刻 t 的累积值, 而 V_t 是未来的支出与收入的差在时刻 t 的贴现值。U_t 由过去的支付与利率决定, 称为后顾式准备金。V_t 由未来的支付与利率决定, 称为前瞻式准备金, 它是在时刻 t 应准备的额度, 以满足未来义务。后顾式准备金与前瞻式准备金的差是合同在时刻 t 时的当前价值。

下面给出准备金的几种表示。

定理 1 由 (42) 式定义的后顾式准备金 U_t 满足下面的向前微分方程

$$\mathrm{d}U_t = U_t r_t \mathrm{d}t + \mathrm{d}A_t \tag{44}$$

上述方程的初始条件为

$$U_0 = A_0 \tag{45}$$

另一方面, U_t 的积分形式为

$$U_t = A_t + \int_0^t U_\tau r_\tau \mathrm{d}\tau \tag{46}$$

U_t 的解析解为

$$U_t = A_t + \int_0^t e^{\int_\tau^t r} A_\tau r_\tau \mathrm{d}\tau \tag{47}$$

$$U_t = e^{\int_0^t r}A_t - \int_0^t e^{\int_\tau^t r}(A_t - A_\tau) r_\tau \mathrm{d}\tau \tag{48}$$

证明 把 (42) 式视为 $U_t = X_t Y_t$, 其中 $X_t = \mathrm{e}^{\int_0^t r}, Y_t = \int_{0-}^{t} \mathrm{e}^{-\int_0^\tau r}\mathrm{d}A_\tau$。应用 §1 的 (37) 式, 可得 (44) 式。初始条件 (45) 由定义即得。对 (44) 式取积分, 即得 (46) 式。

为得到 (47) 式, 从 (42) 式出发, 应用 §1 的 (38) 式, 可得

$$\begin{aligned}\int_{0-}^{t} \mathrm{e}^{-\int_0^\tau r}\mathrm{d}A_\tau &= A_0 + \int_0^t \mathrm{e}^{-\int_0^\tau r}\mathrm{d}A_\tau \\ &= A_0 + \mathrm{e}^{-\int_0^t r}A_t - A_0 - \int_0^t A_\tau \mathrm{e}^{-\int_0^\tau r}(-r_\tau)\mathrm{d}\tau \\ &= \mathrm{e}^{-\int_0^t r}A_t + \int_0^t A_\tau \mathrm{e}^{-\int_0^\tau r} r_\tau \mathrm{d}\tau\end{aligned}$$

再由 (42) 式的定义, 即得 (47) 式。另外, 在上式最后一行中代入关系式 $\mathrm{e}^{-\int_0^t r} = 1 - \int_0^t \mathrm{e}^{-\int_0^\tau r} r_\tau \mathrm{d}\tau$, 再由 (42) 式的定义, 整理后, 可得 (48) 式。

定理 2 前瞻式准备金 V_t 满足下面的向后微分方程

$$\mathrm{d}V_t = V_t r_t \mathrm{d}t - \mathrm{d}B_t \tag{49}$$

上述方程的终值条件为

$$V_T = 0 \tag{50}$$

当 $B_T < \infty$ 时, V_t 的积分形式为

$$V_t = B_T - B_t - \int_t^T V_\tau r_\tau \mathrm{d}\tau \tag{51}$$

V_t 的解析解为

$$V_t = B_T - B_t - \int_t^T \mathrm{e}^{-\int_t^\tau r}(B_T - B_\tau) r_\tau \mathrm{d}\tau \tag{52}$$

$$V_t = \mathrm{e}^{-\int_t^T r}(B_T - B_t) + \int_t^T \mathrm{e}^{-\int_t^\tau r}(B_\tau - B_t) r_\tau \mathrm{d}\tau \tag{53}$$

定理 2 的证明和定理 1 的证明很类似, 这里不再给出。实际上, 定理 1 和定理 2 的结论都有直观解释。例如, (44) 式表明累积值的增加来自当前的利息和新的净收入。同样地, (49) 式表明负债的增加来自当前的利息, 而负债的减少来自已经偿还的部分。(52) 式表明未来的债务在时刻 t 一次偿还时, 它的额度为总债务减去提前偿还所节省的利息。(53) 表明, 债务 V_t 可以在合同终止时偿还, 同时要偿还债务产生的利息。

下面进一步考虑前瞻式准备金 V, 设

$$\mathrm{d}B_t = b_t \mathrm{d}t + \Delta B_t$$

由定理 2 即得如下结论。

定理 3 前瞻式准备金 V 满足下列微分方程

$$\frac{\mathrm{d}}{\mathrm{d}t}V_t = V_t r_t - b_t \tag{54}$$

上述方程在使 r 和 b 连续的时间区间上成立, 而且当 $\Delta B_t \neq 0$ 时, 满足下列粘贴条件

$$V_{t-} = V_t + \Delta B_t \tag{55}$$

当 $T < \infty$ 时, 上述微分方程可从后向前求解, 终值条件为

$$V_{T-} = \Delta B_T \tag{56}$$

§4 寿险前瞻式准备金

1. 一般框架

设被保险人投保年龄为 x, 保险期限为 n 年。以 (x) 表示年龄为 x 的被保险人。当 (x) 在时刻 $t \in (0,n)$ 死亡时, 保额为 b_t, 它在时刻 t 支付。当 (x) 生存至时刻 n 时, 保额 b_n 在时刻 n 支付。在投保时的保费为 π_0, 在 $t \in (0,n)$ 时, 保费以连续方式支付, 支付率为 π_t。另外, 设利率强度为确定性函数 r_t。

在所有的小区间内考虑保额与保费的差, 并贴现至时刻 0 求得的期望现值为

$$-\pi_0 + \int_0^n \mathrm{e}^{-\int_0^\tau r} \ {}_\tau p_x\{\mu_{x+\tau}b_\tau - \pi_\tau\}\mathrm{d}\tau + b_n\mathrm{e}^{-\int_0^n r} \ {}_n p_x \tag{57}$$

在平衡原理下, 须确定保费, 使得 (57) 式为 0。

如果保单在时刻 t 时仍然有效, 前瞻式准备金为

$$V_t = \int_t^n \mathrm{e}^{-\int_t^\tau r} \ {}_{\tau-t}p_{x+t}\{\mu_{x+\tau}b_\tau - \pi_\tau\}\mathrm{d}\tau + b_n\mathrm{e}^{-\int_t^n r} \ {}_{n-t}p_{x+t} \tag{58}$$

把 ${}_{\tau-t}p_{x+t} = \mathrm{e}^{-\int_t^\tau \mu_{x+s}\mathrm{d}s}$ 代入 (58) 式, 即得

$$V_t = \int_t^n \mathrm{e}^{-\int_t^\tau (r_s+\mu_{x+s})\mathrm{d}s}\{\mu_{x+\tau}b_\tau - \pi_\tau\}\mathrm{d}\tau + b_n\mathrm{e}^{-\int_t^n (r_s+\mu_{x+s})\mathrm{d}s} \ {}_{n-t}p_{x+t} \tag{59}$$

比较 (59) 式与 §3 的 (43) 式, 可见 (59) 式可表示为 §3 的 (43) 式的形式, 此时, $\mathrm{d}B_t = (\mu_{x+t}b_t - \pi_t)\mathrm{d}t$, $0 < t < n$, $\Delta B_n = b_n$。

在常见的保险产品中, V_t 取非负值。但是, 在理论上可以构造一些保险产品, 使得 $V_t < 0$ 对某些 t 成立。但在实践中, 这样的产品是不允许的, 这是因为对保单持有人来说, $V_t < 0$ 表示保单持有人的债务, 这就无形中鼓励保单持有人通过退保来取消债务。

2. Thiele 微分方程

由 (59) 式及 §3 的定理 3, 可得下面的结论。

定理 4 V_t 满足如下 Thiele 微分方程

$$\frac{\mathrm{d}}{\mathrm{d}t}V_t = (r_t + \mu_{x+t})V_t + \pi_t - \mu_{x+t}b_t \tag{60}$$

(60) 式对所有的使 r, μ, π, b 连续的 t 成立。而当 $\Delta B_t \neq 0$ 时, 有下列粘贴条件

$$V_{t-} = V_t + \Delta B_t \tag{61}$$

当 $n < \infty$ 时, 上述微分方程可从后向前求解, 终值条件为

$$V_{n-} = \Delta B_n \tag{62}$$

(60) 式的右边给出了每份有效的保单负债的变化率的组成, 利息 r_tV_t 使变化率增大, 来自其他死亡的被保险人的部分 $\mu_{x+t}V_t$, 使变化率增大, 其次就是保费与保额的差。

在平衡原理下, 前瞻式准备金还须满足如下初值条件

$$V_0 = \pi_0 \tag{63}$$

Thiele 微分方程 (60) 式可由直接向后构造法得到。设保单在时刻 $t \in (0, n)$ 时有效。考虑在时间区间 $(t, t + \mathrm{d}t]$ 的各种变化：或者被保险人死亡, 死亡概率为 $\mu_{x+t}\mathrm{d}t + o(\mathrm{d}t)$, 此时 V 的条件期望值为 b_t; 或者被保险人生存, 生存概率为 $1 - \mu_{x+t}\mathrm{d}t + o(\mathrm{d}t)$, 此时 V 的条件期望值为 $-\pi_t\mathrm{d}t + \mathrm{e}^{-r\mathrm{d}t}V_{t+\mathrm{d}t}$。综合起来, 即得

$$V_t = b_t\mu_{x+t}\mathrm{d}t - \pi_t\mathrm{d}t + (1 - \mu_{x+t}\mathrm{d}t)\mathrm{e}^{-r\mathrm{d}t}V_{t+\mathrm{d}t} + o(\mathrm{d}t) \tag{64}$$

(64) 式两边都减去 $V_{t+\mathrm{d}t}$, 然后除以 $\mathrm{d}t$, 再令 $\mathrm{d}t \to 0$, 可得 (60) 式。

3. 储蓄保费与风险保费

把 (60) 式变形, 即得

$$\pi_t = \frac{\mathrm{d}}{\mathrm{d}t}V_t - r_tV_t + (b_t - V_t)\mu_{x+t} \tag{65}$$

(65) 式表明, 在时刻 t 时的保费支付率可以分解为如下两部分

$$\pi_t^s = \frac{\mathrm{d}}{\mathrm{d}t}V_t - r_tV_t \tag{66}$$

$$\pi_t^r = (b_t - V_t)\mu_{x+t} \tag{67}$$

称 π_t^s 为储蓄保费, 它表示为使准备金达到应有的水平, 还需要的超过利息以上的部分; 称 π_t^r 为风险保费, 它表示为履行保险责任, 还需要的超过准备金以上的部分。

4. 从随机过程的观点讨论寿险

以 T_x 表示 (x) 的剩余寿命变量, 引入 $I_t = 1[T_x > t]$, 它表示在时刻 t 前被保险人仍然生存这一事件的示性函数。类似地, 引入 $N_t = 1 - I_t = 1[T_x \leqslant t]$。$I_t$ 和 N_t 都可以视为随机过程, 它们都是右连续的而且存在左极限, 而且 N_t 就是简单的计数过程。

现在把关于单个生命的一些标准的寿险用随机过程表示出来。这些表示的意义在于, 当考虑比较复杂的涉及多个生命以及多种状态的寿险时, 随机过程就成为不可缺少的工具。

(i) 生存保险

$$\mathrm{d}B_t = I_t \mathrm{d}\varepsilon_n(t)$$

其中 B 表示保额, $\varepsilon_n(t) = \begin{cases} 0, & 0 \leqslant t < n \\ 1, & t \geqslant n \end{cases}$。保额的现值为 $\mathrm{e}^{-\int_0^n r} I_n$。

(ii) 定期保险

$$\mathrm{d}B_t = 1_{(0,n]}(t)\mathrm{d}N_t$$

保额的现值为 $\int_0^n \mathrm{e}^{-\int_0^\tau r} \mathrm{d}N_\tau$。

(iii) 生存年金

$$\mathrm{d}B_t = I_t 1_{(0,n)}(t)\mathrm{d}t$$

保额的现值为 $\int_0^n \mathrm{e}^{-\int_0^\tau r} I_\tau \mathrm{d}\tau$。

(iv) 两全保险

$$\mathrm{d}B_t = I_t \mathrm{d}\varepsilon_n(t) + 1_{(0,n]}(t)\mathrm{d}N_t$$

保额的现值为 $\mathrm{e}^{-\int_0^n r} I_n + \int_0^n \mathrm{e}^{-\int_0^\tau r} \mathrm{d}N_\tau$。

为计算以上各种寿险的精算现值, 须用到以下两个关系式

$$\mathbb{E}[I_\tau] = {}_\tau p_x$$

$$\mathbb{E}[\mathrm{d}N_\tau] = {}_\tau p_x \mu_{x+\tau} \mathrm{d}\tau$$

第八章　寿险中的Markov链

为引入寿险中的随机过程模型, 考虑一份在时刻 0 时签发的保单, 如通常的寿险、养老金保险, 或其他的保险如伤残保险、疾病保险等。在寿险中, 给付额与保费通常要视保险合同中规定的被保险人的状态而定, 或者要视状态转移情况而定。为此, 假设存在有限状态空间 $\mathcal{Z}=\{1,\cdots,J\}$, 保单在时刻 0 从状态 1 开始, 在任一时刻保单仅处于一种状态。以 $Z(t)$ 表示保单在时刻 t 的状态, 作为 $[0,\infty)$ 到 $\mathcal{Z}$ 的右连续函数, Z 在任意有限的时间区间内至多存在有限多个跳跃点, 并且 $Z(0)=1$。为考虑保单变化的随机性, 可设 Z 是定义于某个概率空间 $(\Omega,\mathcal{H},\mathbb{P})$ 上的一个随机过程。

在确定概率模型时, 需要考虑两个相互矛盾的因素。一方面, 模型需反映现实问题的基本特征, 而现实问题本身是复杂的, 为此需要建立较为复杂的模型。另一方面, 为了使建立的模型能够容易进行数学处理, 要求建立的模型简单易用。建模的巧妙之处在于, 要在这两者之间达到平衡。

以下讨论倾向于模型的简化, 引入Markov链。在Markov性假设下, 可以很容易计算相关的概率与期望。其次, Markov模型具有普遍适用性, 它能够描述影响保单进展的各种机制。

§1　连续时间Markov链

1. Markov性质

理论上讲, 一个随机过程由它的有限维分布决定。设随机过程 Z 的状态空间

是有限的, 记为 $\mathcal{Z}=\{1,\cdots,J\}$。此时, 有限维分布可由事件 $\bigcap_{q=1}^{r}[Z(t_q)=j_q]$ 的概率确定, 其中 $t_1<\cdots<t_q, j_1,\cdots j_q\in\mathcal{Z}$。因此, 根据条件概率定义,

$$\mathbb{P}[Z(t_q)=j_q,q=1,\cdots,r]=\mathbb{P}[Z(t_r)=j_r\big|Z(t_q)=j_q,q=1,\cdots,r-1]$$
$$\mathbb{P}[Z(t_q)=j_q,q=1,\cdots,r-1]$$

由归纳法可得

$$\mathbb{P}[Z(t_q)=j_q,q=1,\cdots,r]=\prod_{q=1}^{r}\mathbb{P}[Z(t_q)=j_q\big|Z(t_p)=j_p,p=0,\cdots,q-1] \tag{1}$$

在(1)式中, 记 $t_0=0, j_0=1$, 此时$[Z(t_0)=j_0]$ 是必然事件。

假设对于所有的 $t_1<\cdots<t_q$ 和 $j_1,\cdots,j_q\in\mathcal{Z}$,

$$\mathbb{P}[Z(t_q)=j_q\big|Z(t_p)=j_p,p=1,\cdots,q-1]=\mathbb{P}[Z(t_q)=j_q\big|Z(t_{q-1})=j_{q-1}] \tag{2}$$

那么可以得到(1)式的一个简单形式。

可以验证(2)式等价于：对于任意 $t_1<\cdots<t_q<t<t_{q+1}<\cdots<t_r$ 和 $j_1,\cdots,j_q,j,j_{q+1},\cdots,j_r\in\mathcal{Z}$,

$$\begin{aligned}&\mathbb{P}[Z(t_p)=j_p,p=q+1,\cdots,r\big|Z(t)=j,Z(t_p)=j_p,p=1,\cdots,q]\\&=\mathbb{P}[Z(t_p)=j_p,p=q+1,\cdots,r\big|Z(t)=j]\end{aligned} \tag{3}$$

如以 t 表示现在时刻, 那么(3)式表明, 在现在状态给定的情况下, 过程的未来与它的过去是独立的。条件(2)称为 Markov 性。

假设 Z 满足条件(2), 那么称 Z 是状态空间为 $\mathcal{Z}$ 的连续时间 Markov 链, 它的简单转移概率记为

$$p_{jk}(t,u)=\mathbb{P}[Z(u)=k|Z(t)=j] \tag{4}$$

其中 $t<u$ 且 $j,k\in\mathcal{Z}$。如果(2)式成立, 那么(1)式可简化为

$$\mathbb{P}[Z(t_q)=j_q,q=1,\cdots,r]=\prod_{q=1}^{p}p_{j_{q-1}j_q}(t_{q-1},t_q) \tag{5}$$

由简单转移概率可得从 j 到某个子集 $\mathcal{K}\subset\mathcal{Z}$ 的转移概率如下

$$p_{j\mathcal{K}}(t,u)=\mathbb{P}[Z(u)\in\mathcal{K}\big|Z(t)=j]=\sum_{k\in\mathcal{K}}p_{jk}(t,u) \tag{6}$$

显然,

$$p_{j\mathcal{Z}}(t,u)=\sum_{k\in\mathcal{Z}}p_{jk}(t,u)=1 \tag{7}$$

2. Markov 性质的另一个定义

现在简要地概述关于 Markov 性的更一般的定义。对于 $\mathcal{T} \subset [0,\infty)$，设 $\mathcal{H}_{\mathcal{T}}$ 表示由 $\{Z(t)\}_{t\in\mathcal{T}}$ 产生的所有事件构成的集合，它表示在时间集 $\mathcal{T}$ 内能够观察到的关于 Z 的所有信息。例如，$\mathcal{H}_{\{t\}}$ 是过程在时刻 t 的信息，它由基本事件 $\varnothing,\Omega$ 和 $[Z(t)=j], j=1,\cdots,J$ 以及这些事件的并集组成。更一般地，$\mathcal{H}_{\{t_1,\cdots,t_p\}}$ 是过程在时刻 $t_1,\cdots,t_p$ 的信息。简记 $\mathcal{H}_t = \mathcal{H}_{\leqslant t} = \mathcal{H}_{[0,t]}$（过程在时刻 t 之前的所有历史信息，而且包括在时刻 t 的信息），$\mathcal{H}_{<t} = \mathcal{H}_{[0,t)}$（过程在时刻 t 之前的所有历史信息，但不包括时刻 t 的信息），$\mathcal{H}_{>t} = \mathcal{H}_{(t,n]}$（过程在时刻 t 之后的未来信息）。

如果对于任意的 $B \in \mathcal{H}_{>t}$，下式成立

$$\mathbb{P}[B|\mathcal{H}_{\leqslant t}] = \mathbb{P}[B|\mathcal{H}_{\{t\}}] \tag{8}$$

那么称过程 Z 为 Markov 过程。(8)式是(3)式的一般形式。

Markov 性的另一个定义是：对于任意的 $A \in \mathcal{H}_{<t}$ 和 $B \in \mathcal{H}_{>t}$，下式成立

$$\mathbb{P}[A\cap B|\mathcal{H}_{\{t\}}] = \mathbb{P}[A|\mathcal{H}_{\{t\}}]\mathbb{P}[B|\mathcal{H}_{\{t\}}] \tag{9}$$

即给定过程的现在状态，它的过去与未来是条件独立的。在有限状态空间的情形下，通过有限维分布的表示，容易验证(8)式和(9)式是等价的。此时可取 $A \in \mathcal{H}_{\{t_1,\cdots,t_p\}}$，$B \in \mathcal{H}_{\{t_{p+1},\cdots,t_{p+q}\}}$，其中 $t_1 < \cdots < t_p < t < t_{p+1} < \cdots < t_{p+q}$。

3. Chapman-Kolmogorov 方程

对于给定的 t，事件 $\{Z(t)=j\}, j\in\mathcal{Z}$ 是互不相交的，而且它们的并集是必然事件，从而对 $0\leqslant s\leqslant t\leqslant u$，

$$\begin{aligned}\mathbb{P}[Z(u)=k|Z(s)=i] &= \sum_{j\in\mathcal{Z}}\mathbb{P}[Z(t)=j,Z(u)=k|Z(s)=i]\\ &= \sum_{j\in\mathcal{Z}}\mathbb{P}[Z(t)=j|Z(s)=i]\mathbb{P}[Z(u)=k|Z(s)=i,Z(t)=j]\end{aligned}$$

如果 Z 是 Markov 过程，那么上式简化为

$$p_{ik}(s,u) = \sum_{j\in\mathcal{Z}} p_{ij}(s,t)p_{jk}(t,u) \tag{10}$$

称(10)式为 Chapman-Kolmogorov 方程。

4. 转移强度

原则上，确定一个 Markov 链 Z 等价于确定 $p_{ij}(s,t)$，使得(5)右边定义的概率有一致性。如果 Z 是离散时间 Markov 链，其中 t 的取值范围是一个有限时间集

或可列时间集, $0 = t_0 < t_1 < t_2 < \cdots$, 问题就变得很简单, 对于每个 $j \in \mathcal{Z}$ 和 $p = 1, 2, \cdots$, 只需选择非负数 $p_{jk}(t_{p-1}, t_p)$ 使得 $\sum\limits_{k \in \mathcal{Z}} p_{jk}(t_{p-1}, t_p) = 1$。但对连续时间 Markov 链, 因为在计算(5)式左边的概率时, 不存在最小时间单位, 所以问题就变得很复杂。对连续时间 Markov 链, 一个基本的思路就是光滑性假设, 使得在无穷小时间区间内的转移概率的概念有意义。

更具体地, 对于任意的 $j, k \in \mathcal{Z}, j \neq k$ 和 $t \in [0, \infty)$, 假设以下极限是存在的

$$\mu_{jk}(t) = \lim_{h \downarrow 0} \frac{p_{jk}(t, t+h)}{h} \tag{11}$$

称 $\mu_{jk}(t)$ 为转移强度。进一步假设 $\mu_{jk}(t)$ 是分段连续的。(11)式的另一种表示就是

$$p_{jk}(t, t + \mathrm{d}t) = \mu_{jk}(t)\mathrm{d}t + o(\mathrm{d}t) \tag{12}$$

其中 $o(\mathrm{d}t)$ 是关于 $\mathrm{d}t$ 的无穷小量。(12)式表明, 在一个较短的时间区间里, 转移概率关于区间长度成比例, 比例因子就是转移强度。例如, 假设对于所有的 $k \neq j$ 和 $\tau \in [t, t+1], \mu_{jk}(\tau)$ 近似地为常数, 而其 $\mu_{jk}(\tau) \ll 1$, 那么 $\mu_{jk}(t)$ 是转移概率 $p_{jk}(t, t+1)$ 的估计量。

对于 $j \notin \mathcal{K} \subset \mathcal{Z}$, 定义在时刻 t 时从状态 j 到状态集 $\mathcal{K}$ 的转移强度为

$$\mu_{j\mathcal{K}}(t) = \lim_{u \downarrow t} \frac{p_{j\mathcal{K}}(t, u)}{u - t} = \sum_{k \in \mathcal{K}} \mu_{jk}(t) \tag{13}$$

特别地, 在时刻 t 时从状态 j 转移到其他状态的转移强度为 $\mu_{j, \mathcal{Z} - \{j\}}(t)$, 简记为

$$\mu_{j\cdot}(t) = \sum_{k; k \neq j} \mu_{jk}(t) \tag{14}$$

由(7)式和(12)式, 可得到

$$p_{jj}(t, t + \mathrm{d}t) = 1 - \mu_{j\cdot}(t)\mathrm{d}t + o(\mathrm{d}t) \tag{15}$$

5. Kolmogorov 微分方程

转移概率是关于时间的二元函数, 除非在一些很平凡的情形下, 不可能按某种方式直接确定转移概率。转移强度是关于时间的一元函数, 它很容易理解, 而且是确定模型的很自然的出发点。后面将会看到, 转移强度是模型的基本元素, 它们可唯一地确定转移概率。

假设过程 Z 在时刻 t 处于状态 j, 为得到在未来时刻 u 时过程处于状态 k 的概率, 可考虑在第一个时间区间 $(t, t + \mathrm{d}t]$ 内发生的状况。或者 Z 以概率 $1 - \mu_{j\cdot}(t)\mathrm{d}t$ 停留在状态 j, 在这一事件发生的条件下, Z 在时刻 u 时处于状态 k 的概率为

$p_{jk}(t+\mathrm{d}t,u)$；或者 Z 以概率 $\mu_{jg}(t)$ 转移到另一个状态 g, 在这一事件发生的条件下, Z 在时刻 u 时处于状态 k 的概率为 $p_{gk}(t+\mathrm{d}t,u)$。总之, Z 在时刻 u 时处于状态 k 的概率为

$$p_{jk}(t,u)=(1-\mu_{j\cdot}(t)\mathrm{d}t)p_{jk}(t+\mathrm{d}t,u)+\sum_{g;g\neq j}\mu_{jg}(t)\mathrm{d}t p_{gk}(t+\mathrm{d}t,u)+o(\mathrm{d}t) \tag{16}$$

展开(16)式右边的项, 把 $p_{jk}(t+\mathrm{d}t,u)$ 移至左边, 两边再除以 $\mathrm{d}t$, 最后令 $\mathrm{d}t\downarrow 0$, 可得到如下结论。

定理 1 (Kolmogorov 向后微分方程)　转移概率满足如下微分方程

$$\frac{\partial}{\partial t}p_{jk}(t,u)=\mu_{j\cdot}(t)p_{jk}(t,u)-\sum_{g;g\neq j}\mu_{jg}(t)p_{gk}(t,u) \tag{17}$$

(17)式在 $(0,u)$ 内转移强度的连续点集上成立。上述微分方程的边界条件为

$$p_{jk}(u,u)=\delta_{jk} \tag{18}$$

其中 δ_{jk} 定义为：如果 $j=k$, 那么 $\delta_{jk}=1$；如果 $j\neq k$, 那么 $\delta_{jk}=0$。

对于给定的 k 和 u, 由(18)式和向后微分方程, 按照向后步骤, 可以计算出 $p_{jk}(t,u),t\in[0,u),j=1,\cdots,J$ 的数值。

在 Chapman-Kolmogorov 方程(10)式中, 如果用 $t,t+\mathrm{d}t,j,g$ 代替 s,t,i,j, 就可以直接得到(16)式。定理 1 前面的论证方法在后面会经常用到。

相应地, 也有 Kolmogorov 向前微分方程, 它可以通过考虑时间区间末的状况得到。根据以上推理, 可以得到

$$p_{ij}(s,t+\mathrm{d}t)=\sum_{g;g\neq j}p_{ig}(s,t)\mu_{gj}(t)\mathrm{d}t+p_{ij}(s,t)(1-\mu_{j\cdot}(t)\mathrm{d}t)+o(\mathrm{d}t)$$

根据上式可以得到以下结论。

定理 2 (Kolmogorov 向前微分方程)　转移概率满足如下微分方程

$$\frac{\partial}{\partial t}p_{ij}(s,t)=\sum_{g;g\neq j}p_{ig}(s,t)\mu_{gi}(t)-p_{ij}(s,t)\mu_{j\cdot}(t) \tag{19}$$

(19)式在 (s,∞) 内转移强度的连续点集上成立。上述微分方程的边界条件为

$$p_{jk}(s,s)=\delta_{ij} \tag{20}$$

对于给定的 i 和 s, 通过(20)式和向前微分方程, 按照向前步骤, 可以计算出 $p_{ij}(s,t),t>s,j=1,\cdots,J$ 的数值。

在一些简单的问题中，微分方程有明确的解析解。但在大部分非平凡的问题中，只能应用数值求解方法。

确定了简单转移概率后，可以根据(5)式中的有限维分布计算 $\mathcal{H}_{\{t_1,\cdots,t_r\}}$ 中任意事件的概率。

对一些复杂的事件，它涉及的时间参数较复杂。例如，当 $\mathcal{T}$ 是一个区间时，$\mathcal{H}_{\mathcal{T}}$ 中的事件的概率一般不能用简单转移概率计算。此时，通常可以建立所求概率满足的微分方程，并应用合适的方法求解。

6. 占位概率与似然函数

在一段时间内过程总停留在同一个状态，是个很令人感兴趣的问题。定义从 t 到 u 的时间区间内，过程处于状态 j 的占位概率为

$$p_{\overline{jj}}(t,u) = \mathbb{P}[Z(\tau) = j, \tau \in (t,u] | Z(t) = j] \tag{21}$$

显然，对于 $t \leqslant s \leqslant u, p_{\overline{jj}}(t,u) = p_{\overline{jj}}(t,s)p_{\overline{jj}}(s,u)$，成立。根据向后方程的推导思路与(15)式，可得

$$p_{\overline{jj}}(t,u) = (1 - \mu_{j\cdot}(t)\mathrm{d}t)p_{\overline{jj}}(t + \mathrm{d}t, u) + o(\mathrm{d}t) \tag{22}$$

由(22)式，注意到 $p_{\overline{jj}}(u,u) = 1$，可得

$$p_{\overline{jj}}(t,u) = \mathrm{e}^{-\int_t^u \mu_{j\cdot}} \tag{23}$$

占位概率是一类特殊的初等事件的概率。这里的初等事件描述了过程在某个时间区间内的变化状况。

转移强度决定了 Markov 过程 Z 的概率分布。为说明这一点，考虑 Z 在时间区间 $[0,n]$ 内一条路径。设路径在时刻 $t_0 = 0$ 时处于状态 $g_0 = 1$，从时刻 t_0 到时刻 t_1，它停留在状态 g_0，在 $[t_1, t_1 + \mathrm{d}t_1)$ 内从 g_0 转移到 g_1，然后停留在 g_1，直到时刻 t_2，在 $[t_2, t_2 + \mathrm{d}t_2)$ 内从 g_1 转移到 g_2，如此继续下去，直至在 $[t_{q-1}, t_{q-1} + \mathrm{d}t_{q-1})$ 内从 g_{q-2} 转移到 g_{q-1}，停留在 g_{q-1}，直到时刻 $t_q = n$。这个事件的概率是占位概率和无穷小转移概率的乘积，因此它是转移强度的函数，由(5)式和(12)式可得此概率为

$$\begin{aligned} & p_{\overline{g_0 g_0}}(t_0,t_1)\mu_{g_0 g_1}\mathrm{d}t_1 p_{\overline{g_1 g_1}}(t_1,t_2)\mu_{g_1 g_2}(t_2)\mathrm{d}t_2 \cdots p_{\overline{g_{j_{q-1}} g_{j_q}}}(t_{q-1},t_q) \\ = & \, \mathrm{e}^{\sum\limits_{p=1}^{q-1} \ln \mu_{g_{p-1}g_p}(t_p) - \sum\limits_{p=1}^{q} \sum\limits_{h;h \neq g_{p-1}} \int_{t_{p-1}}^{t_p} \mu_{g_{p-1},h}(s)\mathrm{d}s} \mathrm{d}t_1 \cdots \mathrm{d}t_{q-1} \end{aligned} \tag{24}$$

(24)式右边 $\mathrm{d}t_1 \cdots \mathrm{d}t_{q-1}$ 前面的表达式称为似然函数。

一般地，简单转移概率 $p_{jk}(t,u)$ 不是初等事件的概率。实际上，如果过程能返回到以前访问过的状态，那么从一种状态到另一种状态的可能路径通常是不可数

的。简单转移概率并不像名字所暗示的那样简单。一般地, 它们不能表示为转移强度的具体的表达式, 而只能表示为微分方程的解。

7. 向后和向前积分方程

从向后微分方程出发, 可以得到如下等价的一系列积分方程。首先把(17)式右边第一项移到左边, 然后两边同乘于积分因子 $\mathrm{e}^{\int_t^u \mu_{j\cdot}}$, 得到一个全微分,

$$\frac{\partial}{\partial t}\left(\mathrm{e}^{\int_t^u \mu_{j\cdot}} p_{jk}(t,u)\right) = -\mathrm{e}^{\int_t^u \mu_{j\cdot}} \sum_{g;g\neq j} \mu_{jg}(t) p_{gk}(t,u)$$

上式两边在 $(t,u]$ 上积分并应用(18)式, 可得

$$\delta_{jk} - \mathrm{e}^{\int_t^u \mu_{j\cdot}} p_{jk}(t,u) = -\int_t^u \mathrm{e}^{\int_\tau^u \mu_{j\cdot}} \sum_{g;g\neq j} \mu_{jg}(\tau) p_{gk}(\tau,u)\mathrm{d}\tau$$

最后, 把 δ_{jk} 移到右边, 两边再乘以 $-\mathrm{e}^{-\int_t^u \mu_{j\cdot}}$, 并应用(23)式, 可得到向后积分方程

$$p_{jk}(t,u) = \int_t^u p_{\overline{jj}}(t,\tau) \sum_{g;g\neq j} \mu_{jg}(\tau) p_{gk}(\tau,u)\mathrm{d}\tau + \delta_{jk} p_{\overline{jj}}(t,u) \tag{25}$$

应用同样的方式, 由(19)式可得以下一系列向前积分方程

$$p_{ij}(s,t) = \delta_{ij} p_{\overline{ii}}(s,t) + \sum_{g;g\neq j} \int_s^t p_{ig}(s,\tau)\mu_{gj}(\tau) p_{\overline{jj}}(\tau,t)\mathrm{d}\tau \tag{26}$$

如果把涉及的事件分解为互不相交的初等事件之和, 那么积分方程可以通过对概率求和直接得到。例如, 对于(26)式, 右边第一项表示过程停留在状态 i 的概率, 只有当 $i=j$ 时这一项才出现; 第二项表示在发生中间转移的条件下的概率, 它考虑了所有的状态 $g(g\neq j)$ 和所有时间区间 $(\tau,\tau+\mathrm{d}\tau)\subset(s,t)$, 使得状态 g 在 $(\tau,\tau+\mathrm{d}\tau)$ 内最后一次到达状态 j, 并停留在状态 j。同样地, (25)式右边的第一项, 可通过考虑过程首次离开状态 j 的方式和时间来得到。

下一节将讨论与寿险有关的一些模型。在每个问题中, 根据具体情况, 确定状态与转移强度的名称与符号。

§2　一些例子

1. 只有一种死因的单个生命

以 T_0 表示人的寿命随机变量, T_0 的生存函数记为 $\bar{F}$。此时只有两种状态:“生”与“死”, 分别用 0 和 1 来表示, 那么状态过程 Z 就是

$$Z(t) = 1[T_0 \leqslant t], t \geqslant 0$$

$Z(t)$ 表示到 $t \geqslant 0$ 时的死亡人数。图 1 给出了有两种状态的简单过程, 其中状态 1 为吸收状态。

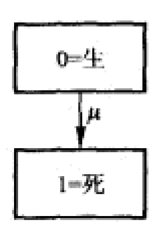


图 1 一种死因的死亡模型

过程 Z 是右连续的, 而且它是 Markov 链, 这是因为如果 $Z(t)=0$, 那么当 $s \leqslant t$ 时, $Z(s)=0$; 如果 $Z(t)=1$, 那么当 $s \geqslant t$ 时,$Z(s)=1$。Z 的转移概率为

$$p_{00}(s,t)=\bar{F}(t)/\bar{F}(s)$$

Chapman-Kolmogorov 方程简化为以下平凡的方程

$$p_{00}(s,u)=p_{00}(s,t)p_{00}(t,u)$$

另外, 唯一的非零转移强度是 $\mu_{01}(t)=\mu(t)$, 而且

$$p_{00}(t,u)=\mathrm{e}^{-\int_t^u \mu} \tag{27}$$

此时,Kolmogorov 微分方程简化为转移强度的定义。

2. 有多种死因的单个生命

前面的过程是非常简单的。在一些复杂的问题中, 随机过程的优势可以充分显示出来。图 2 给出了图 1 的模型的初步扩展, 此时吸收状态有 J 种, 它们对应于多种死因, 例如, "死于事故"、"死于心脏病" 等。转移强度 μ_{0j} 简记为 μ_j。(14)式表明总死亡强度是各个死因的强度之和,

$$\mu(t)=\sum_{j=1}^{J}\mu_j(t) \tag{28}$$

对于一个 t 岁的人, 存活到 u 岁的生存概率 $p_{00}(t,u)$ 由(27)式给出。在这里, 应用扩展的模型, 可以表示各种死因的相对重要程度, 从而能够更好地找出各种具体死因的死亡规律。例如, 简单死亡模型中的 Gompertz-Makeham 法则是两种死因的规律总结, 其中一种死因独立于年龄, 强度为 α, 另一种死因的强度为 $\beta e^{\gamma t}$, 它随着年龄的增长而增大。

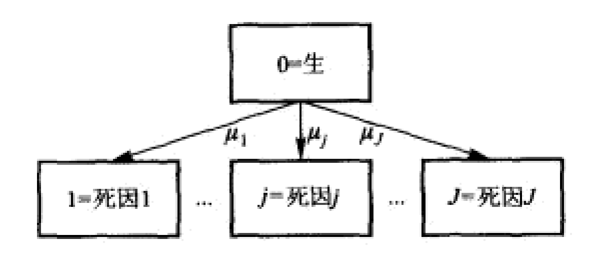


图 2　J种死因的死亡模型

一个 t 岁的人在 u 岁前由于第 j 种死因而死亡的概率为

$$p_{0j}(t,u)=\int_t^u \mathrm{e}^{-\int_t^\tau \mu}\mu_j(\tau)\mathrm{d}\tau \tag{29}$$

上式可以根据(25)式得到, 只需注意到 $p_{\overline{jj}}(t,u)=1$ 即可得到。

由(27)式, (28)式, 以及(29)式, 可以得到如下结论。一种死亡强度 μ_k 的增大会引起生存概率的下降, 同时也引起任何其他死因 $j\neq k$ 的概率下降。因为各种死因概率之和为 1, 因此第 k 种死因的概率会增大。

上述讨论支持了一种观点, 即转移强度是 Markov 过程的基本元素。它们是在保单的每种给定状态下, 对保单产生各种影响的量化表示, 转移概率是这些影响相互作用的结果。

3. 伤残、健康与死亡模型

图 3 给出了一个保险模型, 可分析支付额取决于被保险人健康状况的保险。例如, 疾病保险提供伤残期的年金给付, 另外某些寿险保单允许伤残期内免缴保费。只要重新标识状态, 这个模型也适合于许多其他问题。例如, 在有配偶附加利益的养老金保险中, 状态 a 和 i 可以分别表示“未婚”和“已婚”, 而在失业保险中, 状态 a 和 i 可以分别表示“就业”和“失业”。

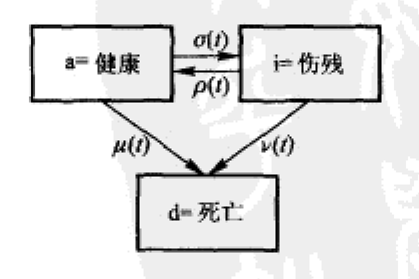


图 3　伤残、健康与死亡的 Markov 链模型

对于一个在时刻 s 健康的人, Kolmogorov 向前微分方程(19)为

$$\frac{\partial}{\partial t}p_{aa}(s,t) = p_{ai}(s,t)\rho(t) - p_{aa}(s,t)(\mu(t)+\sigma(t)) \tag{30}$$

$$\frac{\partial}{\partial t}p_{ai}(s,t) = p_{aa}(s,t)\sigma(t) - p_{ai}(s,t)(\nu(t)+\rho(t)) \tag{31}$$

概率 $p_{ad}(s,t)$ 由 $p_{aa}(s,t)$ 和 $p_{ai}(s,t)$ 决定。初始条件(20)式变为

$$p_{aa}(s,s) = 1 \tag{32}$$

$$p_{ai}(s,s) = 0 \tag{33}$$

对于一个在时刻 s 伤残的人, 向前微分方程是一样的, 只是在所有的概率中, 第一个下标 a 用 i 替换, 并且边界条件为 $p_{ii}(s,s)=1, p_{ia}(s,s)=0$。

当转移强度是比较简单的函数时(如常数函数), 可以得到转移概率的具体解析解。

§3 齐次 Markov 链

1. 矩阵符号

由转移概率可构造如下 $J\times J$ 阶矩阵

$$\mathbf{P}(t,u) = (p_{jk}(t,u)) \tag{34}$$

其中 $p_{jk}(t,u)$ 处于第 j 行第 k 列。特别地,

$$\mathbf{P}(t,t) = \mathbf{I} \tag{35}$$

是 $J\times J$ 阶单位阵。Chapman-Kolmogorov 方程的简化形式为

$$\mathbf{P}(s,u) = \mathbf{P}(s,t)\mathbf{P}(t,u) \tag{36}$$

转移强度矩阵, 又称为无穷小矩阵, 记为

$$\mathbf{M}(t) = (\mu_{jk}(t)) \tag{37}$$

其中 $\mu_{jk}(t)$ 处于第 j 行第 k 列, 而且定义 $\mu_{jj}(t) = -\mu_{j\cdot}(t)$。

应用矩阵形式, Kolmogorov 向后方程(17)式及其边界条件(18)式可以写成

$$\frac{\partial}{\partial t}\mathbf{P}(t,u) = -\mathbf{M}(t)\mathbf{P}(t,u), t\in(0,u); \quad \mathbf{P}(u,u)=\mathbf{I} \tag{38}$$

类似地,Kolmogorov 向前方程(19)式及其附带条件(20)可以写成

$$\frac{\partial}{\partial t}\mathbf{P}(s,t) = \mathbf{P}(s,t)\mathbf{M}(t), t\in(s,\infty); \quad \mathbf{P}(s,s)=\mathbf{I} \tag{39}$$

2. 齐次 Markov 链

如果转移强度不依赖于 t, 即$\mathbf{M}$是常量, 那么称 Markov 链 Z 是齐次的。在这种情况下, 转移概率只取决于时间区间长度, 而与其初始时刻无关。此时, $\mathbf{P}(s,s+t)$ 可简记为 $\mathbf{P}(t)$, Chapman-Kolmogorov 方程(36)式就简化为

$$\mathbf{P}(s+t)=\mathbf{P}(s)\mathbf{P}(t) \tag{40}$$

而(35)式就是 $\mathbf{P}(0)=\mathbf{I}$。由(38)式和(39)式给出的 Kolmogorov 微分方程就简化为同一个向前方程

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t)=\mathbf{P}(t)\mathbf{M}=\mathbf{M}\mathbf{P}(t),t>0; \quad \mathbf{P}(0)=\mathbf{I} \tag{41}$$

由(40)式给出的 Chapman-Kolmogorov 方程, 类似于定义指数函数的泛函方程的矩阵形式。因此, 不难理解(41)式的解为

$$\mathbf{P}(t)=\mathrm{e}^{\mathbf{M}t} \tag{42}$$

其中矩阵指数的定义为：对任意方阵$\mathbf{A}$,

$$\mathrm{e}^{\mathbf{A}}=\sum_{p=0}^{\infty}\frac{1}{p!}\mathbf{A}^p$$

需要指出, 矩阵指数没有完全保留指数函数的所有优良性质。特别地, $\mathrm{e}^{\mathbf{A}+\mathbf{B}}$ 与 $\mathrm{e}^{\mathbf{A}}\mathrm{e}^{\mathbf{B}}$ 通常不相等。容易验证, 如果 $\mathbf{A}$ 和 $\mathbf{B}$ 是可交换的, 即 $\mathbf{AB}=\mathbf{BA}$, 那么 $\mathrm{e}^{\mathbf{A}+\mathbf{B}}=\mathrm{e}^{\mathbf{A}}\mathrm{e}^{\mathbf{B}}$。

为考虑 $\lim\limits_{t\to\infty}\mathbf{P}(t)$ 的存在性, 假设所有状态都是彼此相通的, $\boldsymbol{\pi}'=(\pi_1,\cdots,\pi_J)$ 是平稳分布, 那么 $\lim\limits_{t\to\infty}\mathbf{P}(t)=\mathbf{\Pi}$, 此时

$$\mathbf{\Pi}=\mathbf{1}\boldsymbol{\pi}' \tag{43}$$

其中$\mathbf{1}$是所有的元素都为 1 的 J 维列向量。

在(41)式中令 $t\to\infty$, 由(43)式, 得到 $\mathbf{1}\boldsymbol{\pi}'\mathbf{M}=\mathbf{M}\mathbf{1}\boldsymbol{\pi}\mathbf{i}'=\mathbf{O}$ ($J\times J$ 阶零矩阵), 从而

$$\boldsymbol{\pi}'M=0', \quad \mathbf{M1}=\mathbf{0} \tag{44}$$

其中 $\mathbf{0}$ 是所有元素都为 0 的 J 维列向量。所以, 0 是 $\mathbf{M}$ 的一个特征值, $\boldsymbol{\pi}'$ 和 $\mathbf{1}$ 分别是相应的左特征向量和右特征向量。

现在引用矩阵理论的一些结论。设 $\rho_j, j=1,\cdots,J$ 是 $\mathbf{M}$ 的特征值, 对于每个 j, 设 ϕ_j 是相应的右特征向量。设 $\mathbf{\Phi}$ 是第 j 列为 ϕ_j 的 $J\times J$ 阶矩阵, 那么 $\mathbf{\Phi}^{-1}$ 的第 j 行就是相应于 ρ_j 的左特征向量, 记为 ψ_j'。引入对角阵 $\mathbf{R}(t)=\mathrm{diag}(\mathrm{e}^{\rho_j t})$, 转移矩阵 $\mathbf{P}(t)$ 可表示为

$$\mathbf{P}(t)=\mathbf{\Phi}\mathbf{R}(t)\mathbf{\Phi}^{-1}=\sum_{j=1}^{J}\mathrm{e}^{\rho_j t}\phi_j\psi_j' \tag{45}$$

由(45)式易于计算 $\mathbf{P}(t)$。选择 $\rho_1 = 0, \phi_1 = \mathbf{1}^{J\times 1}, \psi_1' = \boldsymbol{\pi}'$ 可得

$$\mathbf{P}(t) = \mathbf{1}\boldsymbol{\pi}' + \sum_{j=2}^{J} \mathrm{e}^{\rho_j t}\phi_j\psi_j' \tag{46}$$

对转移强度矩阵 $\mathbf{M}$, 所有的 $\rho_j, j = 2, \cdots, J$ 的实部都是负数, 所以(46)式表明转移概率以指数速度收敛于平稳分布。

§4 标准的多状态合同

1. 合同涉及的支付量

当用随机过程描述保单的变化时, 假设 Z 是右连续的、至多存在有限多个间断点, 那么以下定义的示性过程 I_g 和计数过程 N_{gh} 也是右连续的、至多存在有限多个间断点。

$$I_g(t) = 1[Z(t) = g]$$

$I_g(t)$ 表示保单在时刻 t 是否处于状态 g, 它取 1 或 0。

$$N_{gh}(t) = \sharp\{\tau; Z(\tau-) = g, Z(\tau) = h, \tau \in (0, t]\}$$

$N_{gh}(t)$ 表示在时间区间 $(0, t]$ 内发生的从状态 g 转移到状态 $h(h \neq g)$ 的次数。

过程 Z 到达或离开状态 g 时, 相应的 I_g 就增加或减少 1, 这样就把示性过程 $\{I_g(t)\}_{t\geqslant 0}$ 和计数过程 $\{N_{gh}(t)\}_{t\geqslant 0}$ 联系起来。具体地,

$$\mathrm{d}I_g(t) = \mathrm{d}N_{\cdot g}(t) - \mathrm{d}N_{g\cdot}(t), \quad I_g(0) = \delta_{1g} \tag{47}$$

其中下标的点表示加总, 例如,$N_{g\cdot} = \sum_{h;h\neq g} N_{gh}$。

现在考虑标准型保单。保单的支付量函数即合同利益与保费之差, 它有以下形式

$$\mathrm{d}B(t) = \sum_{g} I_g(t)\mathrm{d}B_g(t) + \sum_{g\neq h} b_{gh}(t)\mathrm{d}N_{gh}(t) \tag{48}$$

其中,$\mathrm{d}B_g(t) = b_g(t)\mathrm{d}t + \Delta B_g(t)$, B_g 是保单处于状态 g 时的确定性支付量函数(它是年金的一般形式),b_{gh} 是从状态 g 转移到状态 h 时的确定性支付量函数(它是保险的一般形式)。$\Delta B_g(t)$ 是在时刻 t 的纯生保险。假设函数 b_g 和 b_{gh} 取有限值而且是分段连续的。

在(48)式中, 相对于保险人来说, 支付量取正数表示保额或保险义务, 而保费收入可视为支付量为负。在实际问题中, 保费是以年金的方式缴付的。

假设合同的有限期限是 n。当 $t > n$ 时, 所有的支付函数 $\mathrm{d}B_g(t)$ 和 $b_{gh}(t)$ 都取 0。

2. 现值变量的期望与前瞻准备金

在保单有效期间, 保险人必须始终保持能够满足其未来净负债要求的准备金。在任何时刻 $t \in [0, n]$, 总负债是未来保险义务与保费之差的现值,

$$V(t) = \int_t^n \mathrm{e}^{-\int_t^\tau r} \mathrm{d}B(\tau) \tag{49}$$

数量 $V(t)$ 在时刻 t 时是随机变量, 为得到可行的准备金, 需对 $V(t)$ 进行估计。准备金估计建立在到时刻 t 时已有的信息基础上。在任意时刻 t, 给定保单历史信息 $\mathcal{H}_t$, 保险人应保持的准备金为如下的条件期望

$$V_{\mathcal{H}_t}(t) = \mathbb{E}[V(t)|\mathcal{H}_t] \tag{50}$$

前面假设函数 B_g 和 b_{gh} 都是确定性的, (49)式中的贴现值仅取决于过程 Z 的未来变化。由 Markov 性, (50)式中的条件期望仅依赖于保单的当前状态

$$V_{\mathcal{H}_t}(t) = V_{Z(t)}(t) = \mathbb{E}[V(t)|Z(t)]$$

所以仅需考虑依赖于当前状态的前瞻准备金, 即对 $t \in [0, n], j = 1, \cdots, J$,

$$V_j(t) = \mathbb{E}[V(t)|Z(t) = j] = \mathbb{E}\left[\int_t^n \mathrm{e}^{-\int_t^\tau r} \mathrm{d}B(\tau) | Z(t) = j\right] \tag{51}$$

(51)式是 t 的确定性函数。把(48)式代入(51)式中, 并注意到

$$\mathbb{E}[I_g(\tau)|Z(t) = j] = p_{jg}(t, \tau)$$

$$\mathbb{E}[\mathrm{d}N_{gh}(\tau)|Z(t) = j] = p_{jg}(t, \tau)\mu_{gh}(\tau)\mathrm{d}\tau$$

可得到准备金 $V_j(t)$ 的如下积分表达式

$$V_j(t) = \int_t^n \mathrm{e}^{-\int_t^\tau r} \sum_g p_{jg}(t, \tau) \left(\mathrm{d}B_g(\tau) + \sum_{h;h \neq g} b_{gh}(\tau)\mu_{gh}(\tau)\mathrm{d}\tau\right) \tag{52}$$

上述结果有如下解释: 保单在时刻 τ 处于状态 g 的概率为 $p_{jg}(t, \tau)$, 在时刻 τ 附近, 年金的支付量为 $\mathrm{d}B_g(\tau)$, 此支付量在时刻 t 的期望现值为 $p_{jg}(t, \tau)\mathrm{e}^{-\int_t^\tau r}\mathrm{d}B_g(\tau)$。保单在 $[\tau, \tau + \mathrm{d}\tau]$ 时期从状态 g 转移到状态 h 的概率为 $p_{jg}(t, \tau)\mu_{gh}(\tau)\mathrm{d}\tau$, 此时保险人的支付量为 $b_{gh}(\tau)$, 它在时刻 t 的期望现值为 $p_{jg}(t, \tau)\mu_{gh}(\tau)\mathrm{d}\tau\mathrm{e}^{-\int_t^\tau r}b_{gh}(\tau)$。考虑到未来所有可能的支付时刻和支付类型, 即得到(52)式。

设 $0 \leqslant t < u < n$, 把(52)式右边的积分分解为 $(t, u]$ 和 $(u, n]$ 两个区间上的积分, 对后一积分, 首先应用 Champman-Kolmogorov 方程, 得到

$$\begin{aligned} V_j(t) = & \int_t^u \mathrm{e}^{-\int_t^\tau r} \sum_g p_{jg}(t, \tau) \left(\mathrm{d}B_g(\tau) + \sum_{h;h \neq g} b_{gh}(\tau)\mu_{gh}(\tau)\mathrm{d}\tau\right) \\ & + \mathrm{e}^{-\int_t^u r} \sum_g p_{jg}(t, u)V_g(u) \end{aligned} \tag{53}$$

(53) 式也可以通过考虑保单在时刻 u 所处的状态来得到。

3. 向后(Thiele)微分方程

如果直接应用(52)式，计算的效率不高。对(52)式求导数，并应用 Kolmogorov 向后方程(17)式，可以得到前瞻准备金所满足的微分方程。

定理 3 (Thiele 微分方程的一般形式) (52)式中的依赖于状态的前瞻准备金是以下常微分方程系统的解，

$$\frac{\mathrm{d}}{\mathrm{d}t}V_j(t) = r(t)V_j(t) - b_j(t) - \sum_{k;k\neq j} \mu_{jk}(t)(b_{jk}(t) + V_k(t) - V_j(t)) \tag{54}$$

(54)式对满足以下条件的 t 成立：$\Delta B_g(t) = 0$，而且 $r(t), \mu_{jk}(t), b_j(t)$ 和 $b_{jk}(t)$ 等系数都是连续的。当 $\Delta B_g(t) \neq 0$ 时，粘贴条件为

$$V_j(t-) = \Delta B_j(t) + V_j(t) \tag{55}$$

而(54)式的初始条件为

$$V_j(n-) = \Delta B_j(n) \tag{56}$$

如果沿用 § 1 节第 7 部分的方法，并考虑到上述粘贴条件，可得到如下积分方程

$$V_j(t) = \int_t^u p_{\overline{jj}}(t,\tau)\mathrm{e}^{-\int_t^\tau r}\left(\mathrm{d}B_j(\tau) + \sum_{k;k\neq j} \mu_{jk}(\tau)(b_{jk}(\tau) + V_k(\tau))\mathrm{d}\tau\right) \tag{57}$$

在计算准备金时，最有效的方法就是求微分方程的数值解。注意到微分方程及其边界条件仅涉及模型的基本元素，即 r, μ_{gh} 以及合同中的 B_g, b_g, b_{gh}。而在由(52)式给出的积分表达式涉及转移概率 $p_{gh}(t,u)$，而转移概率一般不存在明显的解析表达式。

关于 V_j 的可微性，这里给出一些说明。在 r, μ_{jk}, b_j 和 b_{jk} 的连续点上，(52)式右边的被积函数是连续的，所以在这些点上，V_j 是连续的。在被积函数的不连续点上，$\frac{\mathrm{d}}{\mathrm{d}t}V_j$ 不存在，但是左导数和右导数是存在的，另外因为至多存在有限多个不连续点，在数值计算中，可以消除在不连续点上的不可微问题。即使不考虑这一点，数值计算程序通常会很精确，它会随着计算过程对系数的变化随时做出调整。唯一比较重要的不可微问题来自年金函数 B_g，它已经由粘贴条件(55)式给出。

顺便指出，粘贴条件可由(53)式得到：用 s, t 替换(53)式中的 t, u。假设 $\Delta B_j(t) \neq 0$，令 $s \to t$，即得粘贴条件(55)式。

注意到 Kolmogorov 向后方程(17)式是 Thiele 方程(54)式的一个特例。转移概率 $p_{jk}(t,u)$ 可以看成是某个简单合同在时刻 t 处于状态 j 的前瞻准备金，该合同

只在时刻 u 的状态为 k 的情况下一次性支付 1, 而且不考虑利息因素。所以, 准备金的数值计算程序同样可应用于转移概率的计算。

定理 3 中的微分方程可由不同的方法得到, 下面简要说明这些方法。

第一种方法是微分法。在(53)式中, 设 $u=t+\mathrm{d}t$, 可以得到一个微分形式, 由此容易推导出(54)式和(55)式。

第二种方法是直接向后论证方法。这种方法已经在定理 1 的 Kolmogorov 向后方程的证明中用到了, 它是一种得到广泛应用的有效工具。这种方法的思路是以在时间区间 $(t,t+\mathrm{d}t]$ 内发生的事件为条件, 其中的转移概率有(12)式给出的形式, 出发点是把(49)式中的积分分解成在 $(t,t+\mathrm{d}t]$ 和 $(t+\mathrm{d}t,n]$ 上的两个积分之和,

$$V(t)=\mathrm{d}E(t)+\mathrm{e}^{-r(t)\mathrm{d}t}V(t+\mathrm{d}t)+o(\mathrm{d}t) \tag{58}$$

假设保单在时刻 t 处于状态 j, 而且在时刻 t, 所有的 $\Delta B_k(t)=0$, 系数 $r(t),\mu_{jk}(t),b_j(t)$ 和 $b_{jk}(t)$ 都是连续的。以小区间 $(t,t+\mathrm{d}t]$ 内发生的事件为条件, 可得到

$$\begin{aligned}V_j(t)=&(1-\mu_{j\cdot}(t)\mathrm{d}t)\left(b_j(t)\mathrm{d}t+\mathrm{e}^{-r(t)\mathrm{d}t}V_j(t+\mathrm{d}t)\right)\\&+\sum_{k;k\neq j}\mu_{jk}(t)\mathrm{d}t\left(b_{jk}(t)+\mathrm{e}^{-r(t)\mathrm{d}t}V_k(t+\mathrm{d}t)\right)+o(\mathrm{d}t)\end{aligned}$$

对上式整理后就得到(54)式, 其中用到了泰勒展开式 $\mathrm{e}^{-r(t)\mathrm{d}t}=1-r(t)\mathrm{d}t+o(\mathrm{d}t)$, $V_g(t+\mathrm{d}t)=V_g(t)+\frac{\mathrm{d}}{\mathrm{d}t}V_g(t)+o(\mathrm{d}t)$, 以及 $(\mathrm{d}t)^2=o(\mathrm{d}t)$。

第三种方法实际上是向后论证方法的严格形式。考虑在时间区间 $(t,n]$ 内是否发生了状态 j 转移到其他状态的事件; 进一步, 如果发生了上述事件, 考虑首次转移的时刻和状态, 可得

$$\begin{aligned}V_j(t)=&\int_t^n \mathrm{e}^{-\int_t^\tau \mu_{j\cdot}}\sum_{k;k\neq j}\mu_{jk}(\tau)\mathrm{d}\tau\left(\int_t^n \mathrm{e}^{-\int_t^\tau r}\mathrm{d}B_j(\tau)+\mathrm{e}^{-\int_t^\tau r}(b_{jk}(\tau)+V_k(\tau))\right)\\&+\mathrm{e}^{-\int_t^n \mu_{j\cdot}}\int_t^n \mathrm{e}^{-\int_t^s r}\mathrm{d}B_j(s)\end{aligned} \tag{59}$$

为了说明(59)式与(57)式是等价的, 只需对(59)式的最后一项积分进行分部积分计算, 它就是

$$\int_t^n \mathrm{e}^{-\int_t^\tau \mu_{j\cdot}}(-\mu_{j\cdot}(\tau)\mathrm{d}\tau)\mathrm{e}^{-\int_t^\tau r}B_j(\tau)+\int_t^n \mathrm{e}^{-\int_t^n \mu_{j\cdot}}\mathrm{e}^{-\int_t^\tau r}\mathrm{d}B_j(\tau)$$

把上式代入(59)式中, 就得到(57)式。再对(57)式取导数, 就得到(54)式。

4. 平衡原理

如果应用平衡原理, 须要求下式成立

$$V_1(0) = -B_1(0) \tag{60}$$

上述条件对合同函数 b_j, B_j 和 b_{jk} 给出了一定的限制, 即对给定的保险义务与保费缴费方式, (60)式最终确定了保费水平。条件(60)式与条件(55)式在性质上是不同的, (55)式直接由前瞻准备金的定义得到。

5. 储蓄保费和风险保费

把方程(54)变形, 可以得到

$$-b_j(t) = \frac{\mathrm{d}}{\mathrm{d}t}V_j(t) - r(t)V_j(t) + \sum_{k;k\neq j} R_{jk}(t)\mu_{jk}(t) \tag{61}$$

其中

$$R_{jk}(t) = b_{jk}(t) + V_k(t) - V_j(t) \tag{62}$$

称 $R_{jk}(t)$ 为在时刻 t 时状态 j 转移到状态 k 的风险总额, 这是因为在发生转移时, 保险人需要立刻支付保险金额 b_{jk}, 另外提供对应于新状态 k 下的准备金, 而保险人已有对应于原来状态 j 下的准备金。因此(61)式的最后一项是当前状态 j 转移到其他状态时, 单位时间内预期净支出, 称为风险保费。

称(61)式右边的前两项之和为储蓄保费, 它表示为保持当前状态下的准备金, 单位时间内的补充量, 它等于准备金的增量与准备金产生的利息之差。

最后, (61)式的左边表示单位时间内支付的保费。所以(61)式表明了保费是如何分解为储蓄部分和风险部分。这是(61)式的一个直观解释。

6. 微分方程的应用

如前所述, 微分方程提供了一种优于其他方法的计算工具, 但是它并不是不可或缺的。但当涉及比较复杂的保险产品的准备金时, 例如当保险合同涉及的函数按一定的方式依赖于准备金时, 微分方程就是不可避免的。最典型的例子就是, 在退保时的偿还额是准备金的一定比例, 另外费用支出也依赖于准备金, 主要的保险利益在某些情况下是准备金的函数等。在这些情形中, 微分方程是确定准备金和保费的不可缺少的工具。下面给出一个例子。

例(寡妇养老金) 一对夫妇购买了一份保单, 它是寿险与丧偶养老金的组合保单。保单规定当夫妇都存活时, 保费的支付率为 c。当丈夫死亡而妻子还生存时, 寡妇养老金支付率为 b; 如果妻子已经去世, 丈夫死亡时的保额为 s 就支付给他们的子女。设保单的期限为 n。图 4 给出了 Markov 模型。

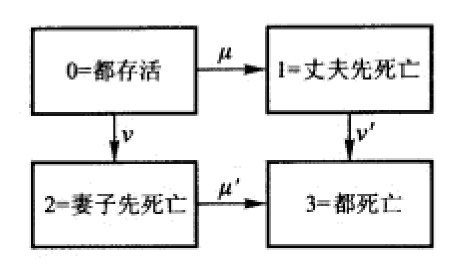


图 4　两个被保险人的模型

假设利率力 r 为常数。在此例中微分方程(54)式的具体形式为(略去平凡方程 $V_3(t)=0$)

$$\frac{\mathrm{d}}{\mathrm{d}t}V_0(t)=\{r+\mu(t)+\nu(t)\}V_0(t)+c-\mu(t)V_1(t)-\nu(t)V_2(t) \tag{63}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V_1(t)=\{r+\nu'(t)\}V_1(t)-b \tag{64}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V_2(t)=\{r+\mu'(t)\}V_2(t)-\mu'(t)s \tag{65}$$

现在考虑一个调整的保险合同, 在时刻 n 前, 如果妻子在丈夫前死亡, 那么准备金的 50% 要支付给丈夫, 这一规定的思路就是没有收到养老金的夫妇应该得到他们的部分储蓄。此时微分方程是必不可少的。在调整的合同中, (63)式右边须减去 $0.5\nu(t)V_0(t)$, 其他方程保持不变, (63)式变为

$$\frac{\mathrm{d}}{\mathrm{d}t}V_0(t)=(r+\mu(t)+0.5\nu(t))V_0(t)+c-\mu(t)V_1(t)-\nu(t)V_2(t) \tag{66}$$

在边界条件 $V_j(n)=0, j=0,1,2$ 下, 可以解出上述方程组。

进一步, 在考虑寡妇养老金问题时, 也可以考虑管理费用支出, 设它部分地依赖于准备金。假设管理费用率在整个区间 $[0,n]$ 内都是当前准备金的一定比例, 设比例为常数 a。此时, 对于每个 j, 在方程(63)式, (64)式, (65)式的右边都减去相应的 $aV_j(t)$ 即可。因此, 管理成本 aV 与利率力减少 a 的影响是相同的。

§5　现值变量的高阶矩

1. 现值变量的矩满足的微分方程

前瞻准备金是未来支付量的贴现值的条件期望, 它是一阶矩。这一节讨论高阶矩。由 Markov 性和(49)式, 只需考虑依赖于状态的条件矩, 对 $q=1,2,\cdots$, 记

$$V_j^{(q)}(t)=\mathbb{E}[V^q(t)|Z(t)=j] \tag{67}$$

定理 4 函数 $V_j^{(q)}$ 是如下微分方程的解

$$\frac{\mathrm{d}}{\mathrm{d}t}V_j^{(q)}(t)=(qr(t)+\mu_{j\cdot}(t))V_j^{(q)}(t)-qb_j(t)V_j^{(q-1)}(t)$$
$$-\sum_{k;k\neq j}\mu_{jk}(t)\sum_{p=0}^{q}\begin{pmatrix}q\\p\end{pmatrix}b_{jk}^p(t)V_k^{(q-1)}(t)$$

上述微分方程在光滑点 t 处成立, 方程的粘贴条件为

$$V_j^{(q)}(t-)=\sum_{p=0}^{q}\begin{pmatrix}q\\p\end{pmatrix}\Delta B_j(t)^pV_j^{(q-p)}(t) \tag{68}$$

而初始条件为

$$V_j^{(q)}(n-)=\Delta B_j^q(n) \tag{69}$$

证明 为了应用向后论证方法, 由(58)式可得

$$V^q(t)=\left(\mathrm{d}B(t)+\mathrm{e}^{-r(t)}V(t+\mathrm{d}t)\right)^q+o(\mathrm{d}t)$$

给定 $Z(t)=j$, 应用二项式展开公式得到

$$V_j^{(q)}(t)=\mathbb{E}\left[\sum_{p=0}^{q}\begin{pmatrix}q\\p\end{pmatrix}(\mathrm{d}B(t))^p\mathrm{e}^{-(q-p)r(t)\mathrm{d}t}V^{q-p}(t+\mathrm{d}t)\Big|Z(t)=j\right]+o(\mathrm{d}t)$$

设在时刻 $t,\Delta B_k(t)=0$, 而且 $r(t),\mu_{jk}(t),b_j(t)$ 和 $b_{jk}(t)$ 都是连续的。以区间 $(t,t+\mathrm{d}t]$ 内发生的事件为条件, 得到

$$\begin{aligned}V_j^{(q)}(t)=&(1-\mu_{j\cdot}(t)\mathrm{d}t)\sum_{p=0}^{q}\begin{pmatrix}q\\p\end{pmatrix}(b_j(t)\mathrm{d}t)^p\mathrm{e}^{-(q-p)r(t)\mathrm{d}t}V_j^{(q-p)}(t+\mathrm{d}t)\\&+\sum_{k;k\neq j}\mu_{jk}(t)\mathrm{d}t\sum_{p=0}^{q}\begin{pmatrix}q\\p\end{pmatrix}b_{jk}^p(t)\mathrm{e}^{-(q-p)r(t)\mathrm{d}t}V_k^{(q-p)}(t+\mathrm{d}t)+o(\mathrm{d}t)\\=&(1-\mu_{j\cdot}(t)\mathrm{d}t)\left(\mathrm{e}^{-qr(t)\mathrm{d}t}V_j^{(q)}(t+\mathrm{d}t)+qb_j(t)\mathrm{d}t\mathrm{e}^{-(q-1)r(t)\mathrm{d}t}V_j^{(q-1)}(t+\mathrm{d}t)\right)\\&+\sum_{k;k\neq j}\mu_{jk}(t)\mathrm{d}t\sum_{p=0}^{q}\begin{pmatrix}q\\p\end{pmatrix}b_{jk}^p(t)V_k^{(q-p)}(t)+o(\mathrm{d}t)\end{aligned}$$

由上式, 进一步推导可得到定理 4 的结论。条件(68)式和(69)式是显然的。

中心矩更直观, 它比原点矩更有用。以 $m_j^{(q)}(t)$ 表示与原点矩 $V_j^{(q)}(t)$ 对应的 q 阶中心矩, 那么

$$m_j^{(q)}(t) = \sum_{p=0}^{q}(-1)^{q-p}\binom{q}{p}V_j^{(p)}(t)\left(V_j^{(1)}(t)\right)^{q-p}, q \geqslant 2 \tag{70}$$

另外记

$$m_j^{(1)}(t) = V_j^{(1)}(t) \tag{71}$$

2. 数值例子

这里要计算与 § 2 节中第 3 部分的“伤残模型”相关的保险的前三阶矩。假设年利率为常数 4.5%, 那么利率力为

$$r = \ln(1.045) = 0.044017$$

另外设不同状态间的转移强度依赖于投保人的年龄x

$$\mu_x = \nu_x = 0.0005 + 0.000075858 \cdot 10^{0.038x}$$

$$\sigma_x = 0.0004 + 0.0000034674 \cdot 10^{0.06x}$$

$$\rho_x = 0.005$$

分别以 0, 1, 2 表示状态 a, i, d。考虑一个 30 岁的男性被保险人, 保险期限为 30 年。因此对 $0 < t < 30$, $\mu_{02}(t) = \mu_{12}(t) = \mu_{30+t} = \nu_{30+t}$, $\mu_{01}(t) = \sigma_{30+t}$, $\mu_{10}(t) = \rho_{30+t}$。(70)式和(71)式定义的中心矩 $m_j^{(q)}(t)$ 在时刻 $t = 0, 6, 12, 18, 24$, 状态 $j = 0, 1$ 的值列在以下各表中。

表 1: 保额为 $1(= b_{02} = b_{12})$的定期保险。

表 2: 健康状态下支付率为 $1(= b_1)$的年金。

表 3: 伤残状态下支付率为 $1(= b_1)$的年金。

表 4: 组合保单, 它包括保额为 $1(= b_{02} = b_{12})$的定期保险, 以及伤残状态下支付率为 $0.5(= b_1)$的年金。该组合保单的净保费支付率为 $0.013108(= -b_0)$, 它在健康状态下支付。

表 1 保额为 1 的定期保险的各阶矩

时刻 t	0	6	12	18	24	30
$m_0^{(1)}(t) = m_1^{(1)}(t)$	0.0683	0.0771	0.0828	0.0801	0.0592	0
$m_0^{(2)}(t) = m_1^{(2)}(t)$	0.0300	0.0389	0.0484	0.0549	0.0484	0
$m_0^{(3)}(t) = m_1^{(3)}(t)$	0.0139	0.0191	0.0262	0.0343	0.0369	0

表 2　健康状态下支付率为 1 的年金的各阶矩

时刻 t	0	6	12	18	24	30
$m_0^{(1)}(t)$	15.763	13.921	11.606	8.698	4.995	0
$m_1^{(1)}(t)$	0.863	0.648	0.431	0.230	0.070	0
$m_0^{(2)}(t)$	5.885	5.665	4.740	2.950	0.833	0
$m_1^{(2)}(t)$	7.795	5.372	3.104	1.290	0.234	0
$m_0^{(3)}(t)$	-51.550	-44.570	-32.020	-15.650	-2.737	0
$m_1^{(3)}(t)$	78.888	49.950	25.099	8.143	0.876	0

表 3　伤残状态下支付率为 1 的年金的各阶矩

时刻 t	0	6	12	18	24	30
$m_0^{(1)}(t)$	0.277	0.293	0.289	0.129	0.119	0
$m_1^{(1)}(t)$	15.176	13.566	11.464	8.708	5.044	0
$m_0^{(2)}(t)$	1.750	1.791	1.646	1.147	0.364	0
$m_1^{(2)}(t)$	11.502	8.987	6.111	3.107	0.716	0
$m_0^{(3)}(t)$	15.960	14.835	11.929	6.601	1.277	0
$m_1^{(3)}(t)$	-101.500	-71.990	-42.500	-17.160	-2.452	0

表 4　组合保单的各阶矩

时刻 t	0	6	12	18	24	30
$m_0^{(1)}(t)$	0.0000	0.0410	0.0751	0.0858	0.0533	0
$m_1^{(1)}(t)$	7.6451	6.8519	5.8091	4.4312	2.5803	0
$m_0^{(2)}(t)$	0.4869	0.5046	0.4746	0.3514	0.1430	0
$m_1^{(2)}(t)$	2.7010	2.0164	1.2764	0.5704	0.0974	0
$m_0^{(3)}(t)$	2.1047	1.9440	1.5563	0.8686	0.1956	0
$m_1^{(3)}(t)$	-12.1200	-8.1340	-4.3960	-1.5100	-0.1430	0

3. 寿险中的偿付能力额度

记 Y 为保单组合的所有未来净负债的现值变量，$m^{(q)}$ 表示 Y 的 q 阶中心矩。Y 的分布的 $1-\varepsilon$ 分位数记为 $y_{1-\varepsilon}$，它的正态幂逼近与前三阶中心矩有关，

$$y_{1-\varepsilon} \approx m^{(1)} + c_{1-\varepsilon}\sqrt{m^{(2)}} + \frac{c_{1-\varepsilon}^2 - 1}{6}\frac{m^{(3)}}{m^{(2)}}$$

其中 $c_{1-\varepsilon}$ 是标准正态分布的 $1-\varepsilon$ 分位数。采用偿付能力控制的准则, $y_{1-\varepsilon}$ 可视为在给定时刻的技术准备金的最低要求。它可分解为保费准备金 $m^{(1)}$ 与波动准备金 $y_{1-\varepsilon}-m^{(1)}$。保单组合风险的一个度量可选为比率值 $R=(y_{1-\varepsilon}-m^{(1)})/P$, 其中 P 是描述保单组合规模的一个适当的度量。作为例子, 考虑 N 份相互独立保单的组合, 每份保单都是表 4 的保单, 而且同时签发。取 P 为每年的总保费收入, 对应于不同的保单数量 N, 签发保单时的 R 值由下表给出

N	10	100	1000	10000	100000
R	48.61	12.00	3.46	1.06	0.332

上述数值结论表明, 当保单数量变大时, 保单组合风险变小。

§6　关于利率的 Markov 链模型

1. 利率力过程

假设经济的变化(或经济中决定利率的因素的变化)以连续时间的齐次 Markov 链 Y 来描述, 状态空间为 $\mathcal{Y}=\{1,\cdots,J^Y\}$, 转移强度为 $\lambda_{ef}, e,f\in\mathcal{Y}, e\neq f$。当经济处于状态 e 时, 对应的利率力为 r_e,

$$r(t)=\sum_e I_e^Y(t)r_e \tag{72}$$

其中 $I_e^Y(t)=1[Y(t)=e]$ 是 Y 在时刻 t 处于状态 e 的示性函数。

图 5 给出了有三种状态的 Markov 链利率力模型, 相应的利率力的取值范围为 $\{0.02,0.05,0.08\}$。利率的转移只能在相邻状态发生, 从任意状态转移出去的总强度为 0.5, 即平均每两年利率力发生一次改变。由对称性, 长时期内平均利率力为 0.05。

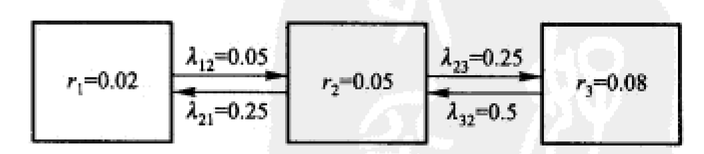


图 5　一个简单的 Markov 链利率力模型

2. 完整的 Markov 模型

引用 § 1 节中关于寿险保单的 Markov 链模型, 并在相应的示性过程和计数过程中给出上标 Z, 以区别于关于利率力的 Markov 链模型 Y。

进一步, 假设过程 Y 和 Z 是独立的, 那么 $X=(Y,Z)$ 是 Markov 链, 状态空间为 $\mathcal{X}=\mathcal{Y}\times\mathcal{Z}$, 转移强度为

$$\mathcal{K}_{ej,fk}(t)=\begin{cases}\lambda_{ef}, & e\neq f, j=k\\ \mu_{jk}(t), & e=f, j\neq k\\ 0, & e\neq f, j\neq k\end{cases}$$

3. 组合模型的矩

考虑 §4 节中的支付流。为了评估合同负债, 在时刻 t 已有信息的条件下, 考虑合同负债的条件分布。这里集中考虑如何确定条件矩, 根据 Markov 性假设, 条件矩仅依赖于当前状态,

$$V_{ej}^{(q)}(t)=\mathbb{E}\left[\left(\int_t^n \mathrm{e}^{-\int_t^\tau r}\mathrm{d}B(\tau)\right)^q \middle| Y(t)=e, Z(t)=j\right]$$

由 §5 节的定理 4, 并对相应符号进行修改, 可以得到如下推论

推论 $V_{ej}^{(q)}(t)$ 是以下微分方程的解

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V_{ej}^{(q)}(t)=&\{qr_e+\mu_{j\cdot}(t)+\lambda_{e\cdot}\}V_{ej}^{(q)}(t)-qb_j(t)V_{ej}^{(q-1)}(t)\\ &-\sum_{k;k\neq j}\mu_{jk}(t)\sum_{p=0}^{q}\binom{q}{p}b_{jk}^p(t)V_{ek}^{(q-p)}-\sum_{f;f\neq e}\lambda_{ef}V_{fj}^{(q)}(t)\end{aligned}\tag{73}$$

(73)式在光滑点 t 上成立, 粘贴条件为

$$V_{ej}^{(q)}(t-)=\sum_{p=0}^{q}\binom{q}{p}(\Delta B_j(t))^p V_{ej}^{(q-p)}(t)\tag{74}$$

对于 $q=2,3,\cdots$, 以 $m_{ej}^{(q)}(t)$ 表示对应于 $V_{ej}^{(q)}(t)$ 的q阶中心矩。而当 $q=1$时, 定义 $m_{ej}^{(1)}(t)=V_{ej}^{(1)}(t)$。计算出原点矩后, 根据

$$m_{ej}^{(q)}(t)=\sum_{p=0}^{q}\binom{q}{p}(-1)^{q-p}V_{ej}^{(p)}(t)\left(V_{ej}^{(1)}(t)\right)^{q-p}$$

可以得到 $q>1$ 阶的中心矩。

4. 组合保单的数值例子

设年龄为 x 的人购买组合保单, 它包括寿险和伤残养老金。记保单签发的时刻为 0, 与保单相关的状态有 $a=$ 健康, $i=$ 伤残, $d=$ 死亡。在时刻 t, 即当被保险人

年龄为 $x+t$ 时，这些状态间的转移强度为

$$\mu_{ad}(t)=\mu_{id}(t)=0.0005+0.000075858\cdot 10^{0.038(x+t)}$$

$$\mu_{ai}(t)=0.0004+0.0000034674\cdot 10^{0.06(x+t)}$$

$$\mu_{ia}(t)=0.005$$

假设利率力可取三个可能值：$r_1=\ln(1.00)=0$(低利率，实际上不考虑利率因素)，$r_2=\ln(1.045)=0.04402$(中等利率)，$r_3=\ln(1.09)=0.08618$(高利率)，这些状态间的转移可用 Markov 链来描述，转移强度矩阵为

$$\Lambda=\lambda\begin{pmatrix}-1 & 1 & 0\\ 0.5 & -1 & 0.5\\ 0 & 1 & -1\end{pmatrix}$$

标度参数 λ 可解释为每单位时间内转移次数的期望，因此 λ 是利率波动性的一个度量。

表 5　前三阶中心矩 $m_{ej}^{(q)}(0), q=1,2,3$

(第二列的 π 值表示初始状态为“中等利率、健康 $(2,a)$”的保单的净保费)

(e,j)			$(1,a)$	$(1,i)$	$(2,a)$	$(2,i)$	$(3,a)$	$(3,i)$
λ	π	q						
0	0.0131	1	0.15	13.39	0.00	7.65	−0.39	5.03
		2	2.55	12.50	0.49	2.70	0.13	0.80
		3	20.45	−99.02	2.11	−12.12	0.37	−2.38
0.05	0.0137	1	0.06	11.31	0.00	7.90	−0.03	5.78
		2	1.61	12.26	0.62	5.41	0.25	2.43
		3	11.94	−42.87	3.20	−4.33	0.94	−0.08
0.5	0.0134	1	0.02	8.43	0.00	7.81	−0.02	7.24
		2	0.65	4.90	0.55	4.15	0.46	3.52
		3	3.34	−13.35	2.59	−10.13	2.02	−7.74
5	0.0132	1	0.00	7.77	0.00	7.70	0.00	7.64
		2	0.51	2.86	0.50	2.91	0.49	2.86
		3	2.26	−12.51	2.20	−12.19	2.14	−11.88
∞	0.0132	1	0.00	7.69	0.00	7.69	0.00	7.69
		2	0.50	2.74	0.50	2.74	0.50	2.74
		3	2.15	−12.37	2.15	−12.37	2.15	−12.37

表 5 给出了在时刻 0 时组合保单现值变量的前三阶中心矩。这里的组合保单的具体内容为：投保年龄 $x=30$, 保单期限 $n=30$, 保险利益包括保额 $1(=b_{ad}=b_{id})$ 的寿险, 以及支付率为 $0.5(=b_i)$ 的伤残养老金, 保费在健康状态下连续支付, 支付率为 $\pi(=-b_a)$, 这里选取的 π 是状态 $(2,a)$ 下的净保费, 这里的 $(2,a)$ 表示"中等利率、健康状况"。

表 5 的前三列构成了一个基准。$\lambda=0$ 表示没有利息波动, 所以得到的结论对应于三个固定利率下的情形。现值变量的二阶矩和三阶矩很强烈地依赖于初始利率力, 实际上, 当利率力增大时, 每行中相应的量的绝对值会变小, 如每行中对应于 $(1,a),(2,a),(3,a)$ 的数值。这是很自然的, 因为利率力的增大会减小贴现因子, 从而未来额度的贴现值就变小。

观察表 5 即知, 随着 λ 值增大, 相对应的三列(如 $(1,a),(2,a),(3,a)$)之间的差异越来越小, 最后完全消失。一个直观解释就是如果利率变化很频繁, 那么初始利率水平的重要性就越来越减弱。

对于与中等利率水平相对应的中间两列, 当 λ 从 0 开始不断增大时, 现值变量的方差在开始时是增大的, 在到达最大值后开始减小, 然后到达稳定值。这一观察结果的直观解释如下：当初始利率为中等利率时, 利率的波动会增大合同最终结果的不确定性, 但是如果利率变化足够频繁, 那么最终结果就与中等水平固定利率下的结果相似。至于第二列的净保费值有类似的变化趋势, 可做同样的解释。

§7 应用鞅方法推导 Thiele 微分方程

在标准的保险保单中, 设状态过程 Z 是 Markov 过程。$\mathcal{H}_t=\sigma\{Z_s;0\leqslant s\leqslant t\}$ 是由 Z 产生的。补偿计数过程 $M_{gh},g\neq h$ 定义为

$$\mathrm{d}M_{gh}(t)=\mathrm{d}N_{gh}(t)-I_g(t)\mu_{gh}(t)\mathrm{d}t \tag{75}$$

$\{M_{gh},g\neq h\}$ 是均值为 0 的平方可积鞅, 而且相互正交, 初值为 0。

设利率过程依赖于状态过程

$$r(t)=r_{Z(t)}=\sum_j I_j(t)r_j \tag{76}$$

为了简便, 以下记 $\int_0^\tau r=\int_0^\tau r(s)\mathrm{d}s$。定义鞅

$$\begin{aligned}M(t)&=\mathbb{E}\left[\int_{0-}^{n}\mathrm{e}^{-\int_0^\tau r}\mathrm{d}B(\tau)\Big|\mathcal{F}_t\right]\\&=\int_{0-}^{t}\mathrm{e}^{-\int_0^\tau r}\mathrm{d}B(\tau)+\mathrm{e}^{-\int_0^t r}\mathbb{E}\left[\int_t^n\mathrm{e}^{-\int_t^\tau r}\mathrm{d}B(\tau)\Big|\mathcal{F}_t\right]\\&=\int_{0-}^{t}\mathrm{e}^{-\int_0^\tau r}\mathrm{d}B(\tau)+\mathrm{e}^{-\int_0^t r}\sum_j I_j(t)V_j(t)\end{aligned} \tag{77}$$

上式最后等式用到了过程 Z 的 Markov 性。

对(77)式应用 Itô 公式, 得到

$$\begin{aligned}\mathrm{d}M(t) =& \mathrm{e}^{-\int_0^t r}\mathrm{d}B(t) + \mathrm{e}^{-\int_0^t r}(-r(t)\mathrm{d}t)\sum_j I_j(t)V_j(t) \\ &+ \mathrm{e}^{-\int_0^t r}\sum_j I_j(t)\mathrm{d}V_j(t) + \mathrm{e}^{-\int_0^t r}\sum_{j\neq k}\mathrm{d}N_{jk}(t)(V_k(t)-V_j(t-))\end{aligned} \tag{78}$$

上式右边最后一项来自 Markov 过程的跳跃点: 当过程在时刻 t 从状态 j 转移到状态 k 时, 状态 j 在 t 之前的准备金立即变为状态 k 在时刻 t 的准备金。因为对应于每个状态的准备金都是确定性函数, 而且是有界变差的, 不连续点至多有可列多个, 所以 Markov 过程在这些不连续点出现跳跃的概率为 0。因此, 不必考虑 $V_j(t)$ 和 $I_j(t)$ 可能的公共不连续点。出于同样的原因, 上式右边最后一项的左极限 $V_j(t-)$ 可改为 $V_j(t)$。

把 $\mathrm{d}B$, $r(t)$ 和 $N_{jk}(t)$ 的表达式代入(78)式中, 可以得到

$$\begin{aligned}\mathrm{d}M(t) =& \mathrm{e}^{-\int_0^t r}\sum_j I_j(t)\left(\mathrm{d}B_j(t) - r_jV_j(t)\mathrm{d}t + \mathrm{d}V_j(t) + \sum_{k;k\neq j}\mu_{jk}(t)\mathrm{d}tR_{jk}(t)\right) \\ &+ \mathrm{e}^{-\int_0^t r}\sum_{j\neq k}R_{jk}(t)\mathrm{d}M_{jk}(t)\end{aligned} \tag{79}$$

其中

$$R_{jk}(t) = b_{jk}(t) + V_k(t) - V_j(t)$$

$R_{jk}(t)$ 表示在时刻 t 从状态 j 转移到状态 k 所导致的风险。

因为(79)式右边是鞅的增量, 所以(79)式右边第一项是两个鞅的增量之差。右边第一项是有界变差的, 后面要说明, 它也是连续的, 所以它是常数。最后可得到

$$\mathrm{d}B_j(t) - r_jV_j(t)\mathrm{d}t + \mathrm{d}V_j(t) + \sum_{k;k\neq j}\mu_{jk}(t)\mathrm{d}tR_{jk}(t) = 0 \tag{80}$$

(80)式就是 Thiele 微分方程。

另外, 由以上推导得到鞅 M 的分解

$$\mathrm{d}M(t) = \mathrm{e}^{-\int_0^t r}\sum_{j\neq k}R_{jk}(t)\mathrm{d}M_{jk}(t)$$

最后说明(80)式是连续函数的增量。含有 $\mathrm{d}t$ 的项是连续的, 而如果 $B_j(t)$ 在时刻 t 是不连续的, 由于 $V_j(t-) = \Delta B_j(t)+V_j(t)$, 所以 $B_j(t-)+V_j(t-) = B_j(t)+V_j(t)$, 即 $B_j(t)+V_j(t)$ 是连续的。

第九章　非寿险中的风险过程

§1　风险过程的破产概念

风险过程的破产理论是非寿险精算理论研究的一个经典专题。现在人们普遍认可 Lundberg 在 1903 年发表的论文是这一理论开始出现的标志。一个世纪来，伴随着一些新的数学分支的出现，尤其是概率论和计算技术的发展，同时由于精算理论研究自身的发展，在破产理论中一些新方法、新结论不断被发现。

首先介绍一下问题的背景。记 $\{U(t), t \geqslant 0\}$ 为保险人的盈余过程。这里所讲的盈余，是指某个初始基金加上保费收入，并减去理赔后剩余部分。这种盈余的定义并非财务意义上的，只是为了数学上处理方便而已。$\{U(t), t \geqslant 0\}$ 可表示如下，

$$U(t) = u + c(t) - S(t)$$

其中 $u \geqslant 0$ 称为初始盈余，$c(t)$ 为时间区间 $(0, t)$ 内的保费收入，$S(t)$ 为时间区间 $(0, t)$ 内的索赔额。自然地，考虑到未来索赔额的不确定性，假设 $S(t)$ 是一个随机过程。在未来某一时刻，盈余过程可能取负值，这种事件形象化地称为破产。一个基本的问题就是对给定的模型与假设，如何计算破产概率。按照时间参数的不同，可以区分连续时间破产概率和离散时间破产概率。

1. 连续时间破产概率

给定盈余过程 $\{U(t), t \geqslant 0\}$，初始盈余为 u，引入 T

$$T = \inf\{t | U(t) < 0, t \geqslant 0\}$$

T 表示破产发生的时刻。进一步定义

$$\psi(u) = \mathbb{P}\{T < \infty\}$$
$$\psi(u,t) = \mathbb{P}\{T < t\}$$

$\psi(u)$ 和 $\psi(u,t)$ 分别称为无限时间破产概率和有限时间破产概率, 很自然地它们具有以下性质。

性质 1 对 $0 \leqslant u_1 \leqslant u_2, 0 \leqslant t_1 \leqslant t_2$,

(i) $\psi(u_2,t) \leqslant \psi(u_1,t)$。

(ii) $\psi(u_2) \leqslant \psi(u_1)$。

(iii) $\psi(u,t_1) \leqslant \psi(u,t_2) \leqslant \psi(u)$。

(iv) $\lim\limits_{t\to\infty} \psi(u,t) = \psi(u)$。

性质 (i) 和 (ii) 表明, 当初始盈余变大时, 破产概率变小。这可以通过对盈余过程样本轨道建立对应关系来验证, 即如初值为 u_2 的样本轨道出现负值, 那么初值为 u_1 的样本轨道也必会出现。反之不然, 即如初值为 u_1 的样本轨道出现负值, 并不能推出初值为 u_2 的样本轨道也会出现。

(iii) 表明, 对同样的初始盈余, 有限时间破产概率随着时间推移会变大, 其上界是无限时间破产概率。该结论的推理类似于 (i) 和 (ii) 中所用到的推理。对某一轨道, 如果它已在 t_1 前出现负值, 那么该轨道在 t_2 前已出现负值。反之不然, 如某一轨道在时间区间 (t_1,t_2) 内首次取负值, 在这种情形下, 该轨道在 $(0,t_2)$ 内取负值, 但在 $(0,t_1)$ 内不取负值。

(iv) 表明, 有限时间破产概率的极限就是无限时间破产概率。该结论似乎显然, 但严格的证明要用到概率公理化定义的单调性。这里不再深入展开。

2. 离散时间破产概率

给定盈余过程 $\{U(t), t = h, 2h, 3h, \cdots\}$, 初始盈余为 u, 时间间隔 $h > 0$, 记

$$T = \inf\{t | U(t) < 0, t = h, 2h, 3h, \cdots\}$$

为破产发生的时刻。引入

$$\psi_h(u) = \mathbb{P}\{T < \infty\}$$
$$\psi_h(u,t) = \mathbb{P}\{T < t\}$$

$\psi(u)$ 和 $\psi(u,t)$ 分别称为无限时间破产概率和有限时间破产概率, 它们满足以下性质。

性质 2 对 $0 \leqslant u_1 \leqslant u_2, 0 \leqslant t_1 \leqslant t_2$,

(i) $\psi_h(u_2, t) \leqslant \psi_h(u_1, t)$。

(ii) $\psi_h(u_2) \leqslant \psi_h(u_1)$。

(iii) $\psi_h(u, t_1) \leqslant \psi_h(u, t_2) \leqslant \psi_h(u)$。

(iv) $\lim\limits_{t \to \infty} \psi_h(u, t) = \psi_h(u)$。

(v) $\psi_h(u, t) \leqslant \psi(u, t), \psi_h(u) \leqslant \psi(u)$。

关于(i)—(iv)的推理与性质 1 中相应的推理完全相同。(v)表明离散时间破产概率比连续时间破产概率要小。这是因为如果离散时间轨道取负值, 从而相应的连续时间轨道取负值。反之不然, 如某一连续轨道在时间区间 $(jh, (j+1)h)$ 取负值, 相应的离散时间轨道不一定取负值。

本章后面的内容主要考虑连续时间无限时间内破产概率, 如无特别说明, 简称为破产概率。

§2 Sparre Andersen 风险模型

1. 模型的定义

在 Sparre Andersen 模型中, 索赔的发生服从一个更新过程, 从而该模型又称为更新风险模型。特别地, 当更新过程是 Poisson 过程, 从而索赔过程是复合 Poisson 过程时, 该更新风险模型称为经典风险模型。

形式上, 如记 T_1 为第一次索赔发生的时刻, T_n 为从第 $n-1$ 次索赔到第 n 次索赔的等待时间。当 $\{T_i, i = 1, 2, \cdots\}$ 是随机变量时, 在时间区间 $(0, t)$ 内的索赔次数是一个随机变量, 记为 $N(t)$。进一步, 当 $\{T_i, i = 1, 2, \cdots\}$ 独立同分布(iid)时, 计数过程 $\{N(t), t \geqslant 0\}$ 就称为更新过程。当 T_i 服从指数分布时, 相应的更新过程称为 Poisson 过程。一般地, 记 T_i 的分布函数和密度函数分别为 K 和 k。

设 $\{X_i, i = 1, 2, \cdots\}$ 是独立同分布(iid)非负随机变量, 其中 X_i 表示第 i 次索赔变量。进一步假设对所有的 $i > 0,\ j > 0, X_i$ 和 T_j 相互独立。记 X_i 的分布函数和密度函数分别为 P 和 p。

设单位时间内的保费收入为 c, Sparre Andersen 风险模型可表示如下,

$$U(t) = u + c(t) - S(t) = u + ct - \sum_{i=1}^{N(t)} X_i$$

特别地, 当 $\{N(t), t \geqslant 0\}$ 是 Poisson 过程时, 对每一 $t \geqslant 0, S(t)$ 是一个复合 Poisson 变量, 从而索赔过程 $\{S(t), t \geqslant 0\}$ 是复合 Poisson 过程。

下面从另外一个观点来看盈余过程。第 n 次索赔发生并支付后, 保险人的盈余变为

$$u + c(T_1 + \cdots + T_n) - (X_1 + \cdots + X_n) = u + \sum_{i=1}^{n}(cT_i - X_i)$$

事实上, 破产只在索赔发生时刻才出现, 破产的发生等价于上述盈余会出现负值。

从实际问题出发, 要求保费收入有一定的保费附加, 因此以下条件是自然的

$$c\mathbb{E}(T_i) > \mathbb{E}(X_i)$$

引入附加因子 θ 如下

$$c\mathbb{E}(T_i) = \mathbb{E}(X_i)(1+\theta)$$

此时 $\psi(u) < 1$。

2. 关于破产概率的 Lundberg 不等式

直观上, Sparre Andersen 风险模型的破产概率大小受两个因素的影响, 即初始盈余值 u 和不依赖于 u 的模型的内在特征, 该内在特征完全由 T_i 和 X_j 的分布确定。下面介绍的 Lundberg 不等式是关于破产概率的经典结论之一。

Lundberg 不等式 假设 $M_X(r) = \mathbb{E}[\mathrm{e}^{rX}]$ 对 $0 \leqslant r < \gamma$ 有定义, 此处 γ 或者为有限数, 或者 $\gamma = \infty$, 而且 $\lim\limits_{r\to\gamma} M_X(r) = \infty$, 那么破产概率满足如下不等式

$$\psi(u) \leqslant \mathrm{e}^{-Ru}$$

其中 R 称为调整系数, 它是如下 Lundberg 基本方程的唯一正解

$$\mathbb{E}[\mathrm{e}^{-(cT_i - X_i)R}] = \mathbb{E}[\mathrm{e}^{-cT_iR}]\mathbb{E}[\mathrm{e}^{X_iR}] = 1$$

证明 在证明不等式前, 首先解释一下 Lundberg 基本方程的来源。对原来的风险模型调整时间参数后, 可以得到另一过程 $\{U(n), n = 0, 1, 2, \cdots\}$ 如下

$$\begin{aligned} &U(0) = u \\ &U(n) = u + \sum_{i=1}^{n}(cT_i - X_i), n \geqslant 1 \end{aligned}$$

显然这里的参数 n 并不表示实际时间参数。进一步, 定义另一过程 $\{Y(n), n = 0, 1, 2, \cdots\}$, 其中 $Y(n) = \mathrm{e}^{-RU(n)}$, 可以验证当 R 是 Lundberg 基本方程的正根时, $\{Y(n), n = 0, 1, 2, \cdots\}$ 是鞅。因此可以应用鞅停时定理来证明上述不等式。

下面给出一种初等证明方法。

(i) 首先假设 R 是存在的, 用归纳法证明不等式。

定义 ${}_n\psi(u), n = 1, 2, \cdots$ 为破产在前 n 次索赔中发生的概率, 那么

$$\psi(u) = \lim_{n\to\infty} {}_n\psi(u)$$

为此只需证明对任意 n, 不等式 ${}_n\psi(u) \leqslant \mathrm{e}^{-Ru}$ 成立。

考虑到第一次索赔发生的时刻, 按破产的定义, 可得

$$\begin{aligned}
{}_1\psi(u) &= \int_0^\infty k(t)\int_{u+ct}^\infty p(x)\mathrm{d}x\mathrm{d}t \\
&\leqslant \int_0^\infty k(t)\int_{u+ct}^\infty \mathrm{e}^{-R(u+ct-x)}p(x)\mathrm{d}x\mathrm{d}t \\
&\leqslant \int_0^\infty k(t)\int_0^\infty \mathrm{e}^{-R(u+ct-x)}p(x)\mathrm{d}x\mathrm{d}t \\
&= \mathrm{e}^{-Ru}\mathbb{E}[\mathrm{e}^{-cT_iR}]\mathbb{E}[\mathrm{e}^{X_iR}] = \mathrm{e}^{-Ru}
\end{aligned}$$

上面最后等式由 Lundberg 基本方程的定义可得。

假设对某 n, 不等式 ${}_n\psi(u) \leqslant \mathrm{e}^{-Ru}$ 成立, 那么通过考虑到第一次索赔是否会导致破产发生, 可得

$$\begin{aligned}
{}_{n+1}\psi(u) &= \int_0^\infty k(t)\int_{u+ct}^\infty p(x)\mathrm{d}x\mathrm{d}t + \int_0^\infty k(t)\int_0^{u+ct} p(x)_n\psi(u+ct-x)\mathrm{d}x\mathrm{d}t \\
&\leqslant \int_0^\infty k(t)\int_{u+ct}^\infty \mathrm{e}^{-R(u+ct-x)}p(x)\mathrm{d}x\mathrm{d}t + \int_0^\infty k(t)\int_0^{u+ct} p(x)\mathrm{e}^{-R(u+ct-x)}\mathrm{d}x\mathrm{d}t \\
&= \int_0^\infty k(t)\int_0^\infty \mathrm{e}^{-R(u+ct-x)}p(x)\mathrm{d}x\mathrm{d}t \\
&= \mathrm{e}^{-Ru}
\end{aligned}$$

(ii) 其次要说明 R 的存在性。下面以经典风险模型为例。对一般的风险模型类似的结论也成立。

设 T_i 服从指数分布, 参数为 λ, 那么 Lundberg 基本方程变为

$$\lambda + cR = \lambda M_X(R)$$

附加因子 θ 在这里的定义为

$$\begin{aligned}
c\mathbb{E}(T_i) &= \mathbb{E}(X_i)(1+\theta) \\
c &= \lambda\mathbb{E}(X_i)(1+\theta)
\end{aligned}$$

所以 Lundberg 基本方程又变形为

$$1 + (1+\theta)\mathbb{E}(X)R = M_X(R)$$

定义函数

$$h(r) = 1 + (1+\theta)\mathbb{E}(X)r - M_X(r)$$

可以验证 $h(0) = 0,\ h'(0) > 0,\ h''(r) = -M_X(r)'' < 0$, 所以 $h(r)$ 是上凸函数。

由假设,$M_X(r)$ 对 $0\leqslant r<\gamma$ 有定义, 此处 γ 或者为有限数, 而且 $\lim\limits_{r\to\gamma}M_X(r)=\infty$, 或者 $\gamma=\infty$, 那么容易验证 $\lim\limits_{r\to\infty}M_X(r)=\infty$ 自动成立。总之,$\lim\limits_{r\to\gamma}h(r)=-\infty$, 从而必然有唯一正数 R 使得 $h\ (R)=0$, R 就是调整系数。

至此证明了 Lundberg 不等式。

例 1　考虑经典风险模型, 设 $P(x)=1-\mathrm{e}^{-\alpha x}$, 那么 Lundberg 基本方程变为

$$1+(1+\theta)\frac{1}{\alpha}R=\frac{\alpha}{\alpha-R}$$

解此方程, 取正根, 得到调整系数为

$$R=\frac{\theta}{1+\theta}\alpha$$

例 2　设 $p(x)=\dfrac{1}{a}x^{-2}\mathrm{e}^{-x}, x\geqslant 1$, 这里 $a=\int_1^\infty x^{-2}\mathrm{e}^{-x}\mathrm{d}x\approx 0.149$, 下面要说明在这种情形下, Lundberg 不等式的条件不成立, 从而调整系数不一定存在。

这是因为

$$M_X(r)=\frac{1}{a}\int_1^\infty \mathrm{e}^{rt}x^{-2}\mathrm{e}^{-x}\mathrm{d}x=\frac{1-r}{a}\int_{1-r}^\infty t^{-2}\mathrm{e}^{-t}\mathrm{d}t$$

由函数 $\Gamma(s)$ 的定义域得知, $M_X(r)$ 只在 $r<1$ 时有定义, 故 $\gamma=1$。而当 $0<r<1$ 时,

$$\begin{aligned}M_X(r)&=\frac{1-r}{a}\int_{1-r}^1 t^{-2}\mathrm{e}^{-t}\mathrm{d}t+\frac{1-r}{a}\int_1^\infty t^{-2}\mathrm{e}^{-t}\mathrm{d}t\\&\leqslant\frac{1-r}{a}\int_{1-r}^1 t^{-2}\mathrm{d}t+\frac{1-r}{a}a=\frac{r}{a}+(1-r)\end{aligned}$$

所以

$$\lim_{r\to 1}M_X(r)\leqslant\frac{1}{a}<\infty$$

§3　应用 Laplace 变换求解经典风险模型的破产概率

1. Laplace 变换

设 $f(x)$ 是定义于 $[0,\infty)$ 上的函数, f 的 Laplace 变换定义为

$$f^*(\xi)=\int_0^\infty \mathrm{e}^{-\xi x}f(x)\mathrm{d}x$$

如果对 $\xi\in(a,b)$, Laplace 变换 $f^*(\xi)$ 存在, 那么 $f^*(\xi)$ 唯一确定了 $f(x)$。下面给出一些特殊的函数的 Laplace 变换。

$f(x)$	$f^*(\xi)$	$f^*(\xi)$的定义域
$e^{-\alpha x}$	$1/(\xi+\alpha)$	$\xi > -\alpha$
$e^{\alpha x}$	$1/(\xi-\alpha)$	$\xi > \alpha$
1	$1/\xi$	$\xi > 0$
x	$1/\xi^2$	$\xi > 0$
x^n	$\Gamma(n+1)/\xi^{n+1}$	$\xi > 0$
$\alpha^2 x e^{-\alpha x}$	$[\alpha/(\alpha+\xi)]^2$	$\xi > -\alpha$

Laplace 变换常见的结论如下：

(i) $(f')^*(\xi) = \xi f^*(\xi) - f(0)$

(ii) 记 $\bar{F}(x) = \int_x^\infty f(y)\mathrm{d}y$, 那么

$$\bar{F}^*(\xi) = \frac{1-f^*(\xi)}{\xi}$$

(iii) 设 $h(x) = \int_0^x f(y)g(x-y)\mathrm{d}y$, 那么 $h^*(\xi) = f^*(\xi)g^*(\xi)$。

2. 应用 Laplace 变换求解破产概率

为应用 Laplace 变换求解破产概率 $\psi(u)$, 考虑在 $[0,\mathrm{d}t]$ 内盈余过程的变化。在此区间内, 没有索赔发生的概率为 $1-\lambda\mathrm{d}t$, 盈余过程的值到达 $u+c\mathrm{d}t$; 有索赔发生的概率为 $\lambda\mathrm{d}t$, 而且只有索赔额较大时, 即当 $X > u+c\mathrm{d}t$ 时, 破产才发生, 当 $X \leqslant u+c\mathrm{d}t$ 时, 破产不发生。

从而

$$\begin{aligned}\psi(u) =&(1-\lambda\mathrm{d}t)\psi(u+c\mathrm{d}t)\\ &+\lambda\mathrm{d}t\left\{\int_0^{u+c\mathrm{d}t}\psi(u+c\mathrm{d}t-x)f_X(x)\mathrm{d}x + 1\times[1-F_X(u+c\mathrm{d}t)]\right\}\end{aligned}$$

由上式可得,$\psi(u)$ 满足如下积分微分方程

$$c\psi'(u)+\lambda\left\{\int_0^u \psi(u-x)f_X(x)\mathrm{d}x + 1 - F_X(u) - \psi(u)\right\} = 0$$

对上式两边取 Laplace 变换, 得到

$$c\xi\psi^*(\xi) - c\psi(0) + \lambda[\psi^*(\xi)f^*(\xi) + \bar{F}^*(\xi) - \psi^*(\xi)] = 0$$

整理得到

$$\begin{aligned}\psi^*(\xi) &= \frac{c\psi(0)-\lambda\bar{F}^*(\xi)}{c\xi+\lambda(f^*(\xi)-1)} = \frac{1}{\xi}\frac{\psi(0)-(\lambda/c)\bar{F}^*(\xi)}{1+(\lambda/c)[(f^*(\xi)-1)/\xi]}\\ &= \frac{1}{\xi}\frac{\psi(0)-(\lambda/c)\bar{F}^*(\xi)}{1-(\lambda/c)\bar{F}^*(\xi)}\end{aligned}$$

如记 $c = \lambda\mu(1+\theta)$, 那么

$$\psi^*(\xi) = \frac{1}{\xi}\frac{\psi(0)(1+\theta) - (1/\mu)\bar{F}^*(\xi)}{1+\theta-(1/\mu)\bar{F}^*(\xi)}$$

当 $\xi > 0$ 时, 上述表示的分母是大于零的, 这是因为

$$\frac{\bar{F}^*(\xi)}{\mu} = \int_0^\infty \mathrm{e}^{-\xi x}\frac{1-F_X(x)}{\mu}\mathrm{d}x \leqslant \int_0^\infty \frac{1-F_X(x)}{\mu}\mathrm{d}x = 1$$

因此当 $\xi \downarrow 0$ 时, 为使 $\psi^*(\xi)$ 的极限存在, 要求

$$\lim_{\xi\downarrow 0}\frac{\bar{F}^*(\xi)}{\mu} = \psi(0)(1+\theta)$$

因为 $\dfrac{\bar{F}^*(\xi)}{\mu}$ 是一个密度函数的 Laplace 变换, 所以上式左边为 1。最后得到

$$\psi(0) = \frac{1}{1+\theta}$$

把 $\bar{F}^*(\xi)$ 和 $\psi(0)$ 的表达式代入前面 $\psi^*(\xi)$ 的表达式, 最后得到

$$\psi^*(\xi) = \frac{1}{\xi}\frac{\xi-(1/\mu)(1-f^*(\xi))}{(1+\theta)\xi-(1/\mu)(1-f^*(\xi))}$$

对一些特殊的索赔变量,$f^*(\xi)$ 是有理函数. 那么 $\psi^*(\xi)$ 也是有理函数, 通过反演 Laplace 变换, 可求 $\psi(u)$。

顺便指出, 设 $F(x)$ 为某正随机变量的概率分布函数, 期望为 μ, 那么 $\dfrac{\bar{F}(x)}{\mu}$ 是概率密度函数, 相应的分布称为 F 的均衡分布。

§4 索赔变量服从 Phase 分布时经典风险模型的破产概率

本节有两个目的: (ⅰ)介绍 Phase 分布的概念; (ⅱ)介绍在经典风险模型中, 当索赔变量服从 Phase 分布时, 破产概率的矩阵表示。为简便计, 本节仅考虑经典风险模型, 实际上本节的内容可以在 Sparre Andersen 风险模型中统一讨论。这些内容是破产理论中较新的内容。至于本节所引用结论的证明, 下面不再给出。

1. Phase 分布

考虑一个连续时间 Markov 链 $\{X_t, t \geqslant 0\}$, 状态空间由 1 个吸收状态和 m 个非常返状态构成, 可记为 $\{0,1,\cdots,m\}$。X_t 表示在时刻 t 时 Markov 链所处的状态。向量 $\boldsymbol{\alpha} = (\alpha_1,\alpha_2,\cdots,\alpha_m)$表示过程初始状态为 m 个非常返状态的概率, 相应

地 $\boldsymbol{\alpha}_0$ 表示过程初始状态为吸收状态的概率。在以下所有讨论中, 假设 $\boldsymbol{\alpha}_0=\mathbf{0}$。记矩阵 $\boldsymbol{Q}$ 为连续时间 Markov 链的转移强度矩阵,

$$\boldsymbol{Q}=\begin{bmatrix}0 & 0\\ \boldsymbol{s}_0^{\mathrm{T}} & \boldsymbol{S}\end{bmatrix}$$

其中 $\boldsymbol{S}=[s_{ij}]$ 表示非常返状态之间的转移强度,$\boldsymbol{s}_0^{\mathrm{T}}=[s_{i0}]$ 是列向量, 它表示从非常返状态到吸收状态的转移强度。由转移强度矩阵的定义, 即得 $\boldsymbol{s}_0^{\mathrm{T}}=-S\boldsymbol{e}^{\mathrm{T}}$, 其中 $\boldsymbol{e}^{\mathrm{T}}$ 是列向量, 它的每个元素都为 1。最后, 记 V 为过程到达吸收状态的时间变量, 它的密度函数 $f(x)$ 有以下形式

$$f(x)=\boldsymbol{\alpha}\exp(x\boldsymbol{S})\boldsymbol{s}_0^{\mathrm{T}},x>0$$

由此可得 V 的分布函数 $F(x)$ 以及 V 的期望

$$F(x)=1-\boldsymbol{\alpha}\exp(x\boldsymbol{S})\boldsymbol{e}^{\mathrm{T}}$$

$$\mathbb{E}[V]=-\alpha S^{-1}\boldsymbol{e}^{\mathrm{T}}$$

直观地, 在 $f(x)$ 表达式中出现的 $\boldsymbol{\alpha}\exp(x\boldsymbol{S})$ 可理解如下

$$\boldsymbol{\alpha}\exp(x\boldsymbol{S})=\mathbb{P}(V>x,X_x=j;j=1,2,\cdots,m)$$

即 $\boldsymbol{\alpha}\exp(x\boldsymbol{S})$ 中的第 j 个元素的含义就是: 到时刻 x 时吸收事件还未发生, 而且 Markov 链在时刻 x 的状态为 j 的概率。

上述类型的变量 V 或分布 F, 由两组参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{S}$ 确定, 称为 Phase 变量或分布。对这类分布, F的均衡分布也是 Phase 分布, 参数为 $\boldsymbol{\pi}$ 和 $\boldsymbol{S}$, 其中

$$\boldsymbol{\pi}=-\boldsymbol{\alpha}\boldsymbol{S}^{-1}/\mathbb{E}[V]$$

例 3 设变量 X 的分布是 Phase 分布, 参数为 $\alpha=(0.2,0.3,0.4,0.1)$,

$$S=\begin{bmatrix}-1 & 1 & 0 & 0\\ 0 & -3 & 3 & 0\\ 0 & 0 & -2 & 2\\ 0 & 0 & 0 & -4\end{bmatrix}$$

那么 X 的密度函数有如下形式

$$f(x)=\frac{4}{5}\mathrm{e}^{-x}+\frac{14}{5}\mathrm{e}^{-2x}-\frac{24}{5}\mathrm{e}^{-3x}+\frac{8}{5}\mathrm{e}^{-4x},x>0$$

2. 经典风险模型中破产概率的矩阵表示

在经典风险模型通常的假设下, 关于破产概率有以下结论。

定理 在经典风险模型中, 设初始盈余值为 $u \geqslant 0$, 索赔变量是 Phase 变量, 参数为 $\boldsymbol{\alpha}$ 和 $\boldsymbol{S}$, 那么破产概率有以下矩阵表示

$$\psi(u) = \boldsymbol{\alpha}_+ \exp(uB)\boldsymbol{e}^{\mathrm{T}}$$

其中 $\boldsymbol{B} = \boldsymbol{S} + \boldsymbol{D}, \boldsymbol{D} = \boldsymbol{s}_0^{\mathrm{T}}\boldsymbol{\alpha}_+, \boldsymbol{\alpha}_+ = \boldsymbol{\rho\eta}, \boldsymbol{\rho} = \psi(0) = 1/(1+\theta), \eta = -\boldsymbol{\alpha}\boldsymbol{S}^{-1}/\mu, \mu = \mathbb{E}[X_i]$。

以下数值例子都可以应用数学软件(如 Mathematica)直接求解。

例 4 设索赔变量是两个指数分布变量的混合, 指数分布参数分别为 3 和 7, 权重都是 1/2。设 Poisson 参数 $\lambda = 1$, 保费附加因子 $\theta = 0.4$。此时 $\rho = 5/7, \boldsymbol{\alpha} = (1/2, 1/2), \boldsymbol{\alpha}_+ = (1/2, 3/14)$,

$$\boldsymbol{S} = \begin{bmatrix} -3 & 0 \\ 0 & -7 \end{bmatrix}, \boldsymbol{D} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} [1/2 \quad 3/14], \boldsymbol{B} = \begin{bmatrix} -3/2 & 9/14 \\ 7/2 & -11/12 \end{bmatrix}$$

此时

$$\exp(u\boldsymbol{B}) = \begin{bmatrix} \frac{9}{10}\mathrm{e}^{-u} + \frac{1}{10}\mathrm{e}^{-6u}, & \frac{9}{70}\mathrm{e}^{-u} - \frac{9}{70}\mathrm{e}^{-6u} \\ \frac{7}{10}\mathrm{e}^{-u} - \frac{7}{10}\mathrm{e}^{-6u}, & \frac{1}{10}\mathrm{e}^{-u} + \frac{9}{10}\mathrm{e}^{-6u} \end{bmatrix}$$

最后得到

$$\psi(u) = (24\mathrm{e}^{-u} + \mathrm{e}^{-6u})/35$$

例 5 设索赔变量 X 服从 Erlang(3)分布, 即 X 可表示为三个独立同分布的指数变量之和,$\mu = 1$。设 Poisson 参数 $\lambda = 1$, 保费附加因子 $\theta = 0.1$, 此时 $\rho = 10/11$, $\boldsymbol{\alpha} = (1, 0, 0)$,

$$\boldsymbol{S} = \begin{bmatrix} -3 & 3 & 0 \\ 0 & -3 & 3 \\ 0 & 0 & -3 \end{bmatrix}$$

因此 $\boldsymbol{\alpha}_+ = (10/33, 10/33, 10/33)$,

$$\boldsymbol{B} = \begin{bmatrix} -3 & 3 & 0 \\ 0 & -3 & 3 \\ 10/11 & 10/11 & -23/11 \end{bmatrix}$$

最后得到

$$\begin{aligned}\psi(u) =& 0.9242\mathrm{e}^{-0.1385u}\\ &-2\{0.0075\cos(1.3807u)+0.0053\sin(1.3807u)\}\mathrm{e}^{-3.9761u}\end{aligned}$$

例 6 设索赔变量 X 的密度函数 $p(x)$ 为

$$p(x)=\sum_{i=1}^{2}\sum_{j=1}^{3}q_{ij}\frac{\beta_i^j x^{j-1}\mathrm{e}^{-\beta_i x}}{(j-1)!}, x>0$$

其中 $\beta_1=1$, $\beta_2=0.5, q_{11}=0.05$, $q_{12}=0.1$, $q_{13}=0.5$, $q_{21}=0.05$, $q_{22}=0.2$, $q_{23}=0.1$。如表示为 Phase 分布形式, 那么 $\alpha=(0.5,0.1,0.05,0.1,0.2,0.05)$,

$$\boldsymbol{S}=\begin{bmatrix}-1 & 1 & 0 & 0 & 0 & 0\\ 0 & -1 & 1 & 0 & 0 & 0\\ 0 & 0 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & -0.5 & 0.5 & 0\\ 0 & 0 & 0 & 0 & -0.5 & 0.5\\ 0 & 0 & 0 & 0 & 0 & -0.5\end{bmatrix}$$

设 Poisson 参数 $\lambda=1$, 保费附加因子 $\theta=1$, 那么

$$\boldsymbol{\alpha}_+=(0.07692,0.09231,0.10000,0.03077,0.09231,0.10769)$$

$$\boldsymbol{B}=\begin{bmatrix}-1.00000 & 1.00000 & 0 & 0 & 0 & 0\\ 0 & -1.00000 & 1.00000 & 0 & 0 & 0\\ 0.07692 & 0.09231 & -0.90000 & 0.03077 & 0.09231 & 0.10769\\ 0 & 0 & 0 & -0.50000 & 0.50000 & 0\\ 0 & 0 & 0 & 0 & -0.50000 & 0.50000\\ 0.03846 & 0.04615 & 0.05000 & 0.01539 & 0.04615 & -0.44615\end{bmatrix}$$

最后得到

$$\begin{aligned}\psi(u) =& 0.01518\mathrm{e}^{-0.6154u}+0.51614\mathrm{e}^{-0.2005u}\\ &+0.03296\mathrm{e}^{-0.5495u}\cos(0.14324u)-0.06429\mathrm{e}^{-1.2157u}\cos(0.30563u)\\ &-0.06059\mathrm{e}^{-0.5495u}\sin(0.14324u)-0.05064\mathrm{e}^{-1.2157u}\sin(0.30563u)\end{aligned}$$

§5 鞅方法在非寿险定价中的应用

1. 引言

随机过程理论中的鞅可以帮助理解非寿险保单的定价。特别地，本节中涉及鞅、Brown 运动以及停时等概念。

考虑非寿险保单，如汽车综合险，软件供应商的职业补偿保险等，假设投保按年度进行。一个很自然的问题就是：在理论上，如何计算年度保费？以 P_n 表示第 n 年的保费。

上述问题的出发点就是每年内发生的索赔。因为索赔总额是不确定的，所以可视为随机变量，以 S_n 表示第 n 年内的总索赔额。此外还有一些其他考虑，例如，对于较大的风险，还应考虑用来支持风险的资本金，以 U 表示首期资本金。在实践中，还应考虑到很多其他因素，如竞争者的费用支出及保险费率。为了方便起见，这里忽略其他因素。

假设随机变量 $\{S_n\}_{n=1}^{\infty}$ 是相互独立的，但不必同分布。为了能够计算 P_n，需要计算或者至少能够估计 S_n 的分布。如果知道第 n 年内的索赔次数与单个索赔额的分布，可以应用著名的 Panjer 递推公式来计算 S_n 的分布。在某些条件下，如 S_n 是大量独立索赔之和时，可以假设 S_n 近似服从正态分布。在以下讨论中，有时假设

$$S_n \sim N(\mu_n, \sigma_n^2) \tag{1}$$

其中 μ_n, σ_n 是参数。

2. 标准差原理

按照标准差原理，确定保费时，应使得总索赔额超过保费收入的概率足够小，如 $1-p$。用公式来描述，就是

$$\mathbb{P}\{S_n < P_n\} = p \tag{2}$$

如果 S_n 服从(1)式给出的正态分布，那么容易得到

$$P_n = \mu_n + \gamma_p \sigma_n \tag{3}$$

其中 γ_p 满足 $\Phi(\gamma_p) = p$, $\Phi(\cdot)$ 是标准正态分布的分布函数。

公式(3)表明，在第 n 年的 P_n 应该等于总索赔额的平均值加上总索赔额标准差的一定比例。注意到，比例因子 γ_p 不依赖于 n。在精算学中，称公式(3)为计算保费的标准差原理，γ_p 称为附加因子。

3. 效用函数与方差原理

为了介绍方差原理，首先引入效用函数的概念。假设金钱的效用由效用函数 $u(\cdot)$ 来表示。直观上，$u(x)$ 描述了x数量的金钱的价值。为便于数学处理，要求 u 满足以下两个条件

$$\frac{\mathrm{d}}{\mathrm{d}x}u(x) > 0, \frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) < 0$$

第一个条件表明：金钱越多，价值就越大；第二个条件表明：对相同的金钱额度，价值是递减的。

由效用函数可以通过下式计算 P_n

$$u(W) = \mathbb{E}[u(W + P_n - S_n)] \tag{4}$$

其中，W 是第 n 年的初始财富。(4)式的思路就是：从保险人的角度，如果不承保风险，那么财富的效用值为 $u(W)$；如果承保风险，那么财富在年末的期望效用值为 $\mathbb{E}[u(W + P_n - S_n)]$，保费 P_n 的确定应使得是否承保风险是没有区别的。从这个意义下，保费 P_n 是公平的。

现在假设 $u(\cdot)$ 是指数效用函数，参数为 α ($\alpha > 0$)，

$$u(x) = 1 - \mathrm{e}^{-\alpha x} \tag{5}$$

由(4)和(5)式，可以得到如下形式的 P_n，

$$P_n = \alpha^{-1}\log(M_{S_n}(\alpha)) \tag{6}$$

其中 $M_Z(\cdot)$ 表示随机变量 Z 的矩生成函数，$M_Z(\alpha) = \mathbb{E}[\mathrm{e}^{\alpha Z}]$。注意到，在此特殊情形下，$P_n$ 不依赖于 W。

最后，进一步假设 S_n 服从正态分布，那么

$$M_{S_n}(\alpha) = \exp\{\mu_n\alpha + \frac{1}{2}\alpha^2\sigma_n^2\}$$

从而由(6)式得到

$$P_n = \mu_n + \frac{1}{2}\alpha\sigma_n^2 \tag{7}$$

此时，P_n 是依据方差原理来计算的。

4. 多周期分析 —— 离散时间

公式(3)和(7)给出了在理论上计算 P_n 的两种方法，这两种方法的共同特点在于：

(i) P_n 都是 S_n 的期望值与某正附加项之和。

(ii) 为得到 P_n 的计算公式，第 n 年与其他年度分开考虑。

另一方面, 两者之间的主要区别在于: (3)式中的附加因子 γ_p 有直观意义, 而(6)式或(7)式中的参数 α 的直观解释并不明显。出于这一考虑, 需要综合考虑未来每年末的盈余。

以 U_n 表示第 n 年末的盈余, 定义 $U_0 = U$, 那么

$$U_n = U + \sum_{k=1}^{n}(P_k - S_k), n = 1, 2, 3, \cdots$$

引入以下概率

$$\psi(U) = \mathbb{P}[U_n \leqslant 0, \text{至少存在某}n = 1, 2, 3, \cdots]$$

$\psi(U)$ 表示初始资金在未来能够支持风险的概率, 如用带有感情色彩的话, $\psi(U)$ 表示盈余过程在未来某离散时刻破产的概率。

假设 P_n 由(6)式确定。考虑过程 $\{Y_n\}_{n=0}^{\infty}$, 其中 $Y_n = \exp\{-\alpha U_n\}, n = 0, 1, 2, \cdots$。

首先, $\{Y_n\}_{n=0}^{\infty}$ 关于 $\{S_n\}_{n=1}^{\infty}$ 生成的滤波是鞅。为证明这一结论, 注意到

$$U_{n+1} = U_n + P_{n+1} - S_{n+1}$$

从而

$$\begin{aligned}\mathbb{E}[Y_{n+1}|S_1, \cdots, S_n] &= \mathbb{E}[\exp\{-\alpha(U_n + P_{n+1} - S_{n+1})\}|S_1, \cdots, S_n] \\ &= \mathbb{E}[\exp\{-\alpha(P_{n+1} - S_{n+1})\}|S_1, \cdots, S_n] \exp\{-\alpha U_n\}\end{aligned}$$

因为 $\{S_n\}_{n=1}^{\infty}$ 是一列独立随机变量, 所以

$$\begin{aligned}\mathbb{E}[\exp\{-\alpha(P_{n+1} - S_{n+1})\}|S_1, \cdots, S_n] &= \mathbb{E}[\exp\{-\alpha(P_{n+1} - S_{n+1})\}] \\ &= \exp\{-\alpha P_{n+1}\}\mathbb{E}[\exp\{-\alpha S_{n+1}\}] = 1\end{aligned}$$

上式的最后一步由(6)式得到。因此,

$$\mathbb{E}[Y_{n+1}|S_1, \cdots, S_n] = \exp\{-\alpha U_n\} = Y_n$$

现在应用上述结论来确定 $\psi(U)$ 的上界。为此, 引入正数 $b > U$, 定义如下停时 T

$$T = \min\{n : U_n \leqslant 0\text{或者}U_n \geqslant b\}$$

由可选停时定理可得

$$\mathbb{E}[\exp\{-\alpha U_T\}] = \exp\{-\alpha U\}$$

以 p_b 表示盈余过程 $\{U_n\}$ 在没有到达或超过 b 之前就已经破产的概率, 按全概率公式, 得到

$$\begin{aligned}&\mathbb{E}[\exp\{-\alpha U_T\}] \\ &= \mathbb{E}[\exp\{-\alpha U_T\}|U_T \leqslant 0]p_b + \mathbb{E}[\exp\{-\alpha U_T\}|U_T \geqslant b](1 - p_b) \\ &= \exp\{-\alpha U\}\end{aligned} \tag{8}$$

当 $b \to \infty$ 时, $p_b \to \psi(u)$, 而(8)式中第二项的期望趋于 0, 因此

$$\exp\{-\alpha U\} = \psi(U)\mathbb{E}[\exp\{-\alpha U_T\}|U_T \leqslant 0]$$

从而

$$\psi(U) = \frac{\exp\{-\alpha U\}}{\mathbb{E}[\exp\{-\alpha U_T\}|U_T \leqslant 0]} \leqslant \exp\{-\alpha U\} \tag{9}$$

(9)式给定了破产概率的一个简单上界。另外, 它对如何确定(6)中的参数 α 有参考价值。例如, 设最初的盈余为 10, 如果要求破产概率不超过 1%, 那么 α 满足 $\exp\{-10\alpha\} = 0.01$ 即可, 从而 $\alpha = 0.461$。

最后考虑一种特殊情形。设 $\{S_n\}_{n=1}^{\infty}$ 是独立同分布的随机变量, 而且服从参数为 λ 的复合 Poisson 分布。此时, P_n 不依赖于 n, 即 $P_n = P$, 而且

$$P = \alpha^{-1}\log\mathbb{E}[\exp\{\alpha S_n\}] = \alpha^{-1}\log[\exp\lambda(M_X(\alpha)-1)]$$

其中 $M_X(\cdot)$ 是单个索赔额分布的矩生成函数, 上式就是

$$\lambda + P\alpha = \lambda M_X(\alpha) \tag{10}$$

在上述特殊情形下,(10)式就是破产理论中调整系数所满足的方程, 而 α 就是调整系数。

5. 多周期分析 —— 连续时间

(9)式表明, 如果根据(6)式确定年度保费, 那么离散时间盈余过程的最终破产概率的上界为 $\exp\{-\alpha U\}$。此外, 如果每年内总索赔额服从正态分布, 那么由(6)式可得保险附加与总索赔额变量的方差成比例。

为进一步理解(9)式, 可考虑连续时间模型。在连续时间模型下, 假设在时间区间 $[0,t]$ 内的总索赔额为随机变量 $\mu t + \sigma B(t)$, 其中 $\{B(t)\}_{t\geqslant 0}$ 是标准 Brown 运动。由此即得每年内的总索赔额服从正态分布 $N(\mu,\sigma^2)$, 而且不同年度内的索赔额变量相互独立。Brown 运动具有平稳独立增量性质, 对于任意的 $0 \leqslant s < t$,

$$B(t) - B(s) \sim N((t-s)\mu, (t-s)\sigma^2) \tag{11}$$

另一方面, 假设保费支付率为 P, 它由下式给出

$$P = \mu + \frac{1}{2}\alpha\sigma^2 \tag{12}$$

以 $U(t)$ 表示 t 时的盈余, 那么

$$U(t) = U + Pt - (\mu t + \sigma B(t)) \tag{13}$$

引入符号

$$\psi_c(U) = \mathbb{P}[U(t) \leqslant 0, \text{至少存在某}t > 0]$$

$\psi_c(U)$ 是连续时间盈余过程的最终破产概率。应用前面的鞅方法可求出 $\psi_c(U)$。首先，注意到下式成立

$$\mathbb{E}[\exp\{-\alpha[(P-\mu)t - \sigma B(t)]\}] = 1 \tag{14}$$

(14)式可由(11)式和(12)式，以及正态分布的矩生成函数得到。

定义 $Y(t) = \exp\{-\alpha U(t)\}$，那么过程 $\{Y(t)\}_{t\geqslant 0}$ 关于由 $\{B(t)\}_{t\geqslant 0}$ 生成的滤波是连续鞅。证明如下：对于 $t > s$，

$$\begin{aligned}
&\mathbb{E}[Y(t)|B(u), 0 \leqslant u \leqslant s] \\
&= \mathbb{E}[\exp\{-\alpha U(t)\}|B(u), 0 \leqslant u \leqslant s] \\
&= \mathbb{E}[\exp\{-\alpha(U(t) - U(s))\}\exp\{-\alpha U(s)\}|B(u), 0 \leqslant u \leqslant s] \\
&= \mathbb{E}[\exp\{-\alpha(U(t) - U(s))\}|B(u), 0 \leqslant u \leqslant s]\exp\{-\alpha U(s)\} \\
&= \mathbb{E}[\exp\{-\alpha(U(t) - U(s))\}\exp\{-\alpha U(s)\}
\end{aligned}$$

注意到

$$U(t) - U(s) = (P-\mu) \times (t-s) - \sigma(B(t) - B(s))$$

由平稳增量性，$B(t)-B(s)$ 与 $B(t-s)$ 同分布，从而 $U(t)-U(s)$ 与 $U(t-s)-U$ 同分布，而且

$$\begin{aligned}
&\mathbb{E}[\exp\{-\alpha(U(t) - U(s))\}] \\
&= \mathbb{E}[\exp\{-\alpha(U(t-s) - U)\}] \\
&= \mathbb{E}[\exp\{-\alpha(P-\mu)(t-s) - \sigma(B(t-s))\}] = 1
\end{aligned}$$

上面的最后一式由(14)式得到。总之，以上证明了

$$\mathbb{E}[Y(t)|B(u), 0 \leqslant u \leqslant s] = \exp\{-\alpha U(s)\} = Y(s)$$

即 $\{Y(t)\}_{t\geqslant 0}$ 关于 $\{B(t)\}_{t\geqslant 0}$ 生成的滤波是鞅。

为应用可选停时定理，定义如下停时：对 $b > U$，

$$T_c = \inf\{t : U(t) \leqslant 0\text{或者}U(t) \geqslant b\}$$

由可选停时定理，可得

$$\mathbb{E}[\exp\{-\alpha U(T_c)\}] = \exp\{-\alpha U\}$$

以 p_b 表示盈余过程在到达或超过 b 之前就破产的概率, 那么

$$\begin{aligned}&\mathbb{E}[\exp\{-\alpha U(T_c)\}]\\=&\mathbb{E}[\exp\{-\alpha U(T_c)\}|U(T_c)\leqslant 0]p_b\\&+\mathbb{E}[\exp\{-\alpha U(T_c)\}|U(T_c)\geqslant b](1-p_b)\end{aligned}$$

由于盈余过程的样本路径是连续的, 因此

$$\mathbb{E}[\exp\{-\alpha U(T_c)\}|U(T_c)\geqslant b]=\exp\{-\alpha b\}$$

$$\mathbb{E}[\exp\{-\alpha U(T_c)\}|U(T_c)\leqslant 0]=1$$

最后得到

$$p_b+\exp\{-\alpha b\}(1-p_b)=\exp\{-\alpha U\}$$

如果 $b\to\infty$, 那么 $p_b\to\psi_c(U)$, 由上式, 可得

$$\psi_c(U)=\exp\{-\alpha U\} \tag{15}$$

(15)式就是连续时间盈余过程的 Lundberg 不等式, 实际上, 在以上连续模型中, 它是等式。



第十章　离散时间金融模型

§1　二叉树

最简单的模型包含股票和债券, 股票价格的随机性和货币的时间价值, 在实际的金融市场中是最基本的。

1. 股票

考虑一个时间周期, 它从 $t=0$ 时开始, 到 $t=\delta t$ 时结束。为描述股票价格的不可预期性和随机性, 假设在这段时间内只有两种情况可能发生: 股票价格上升或下降, 用二叉树表示, 见图 1。设股票价格上升的概率为 p, 下降的概率为 $1-p$。在开始时(节点 1)股票的价格为 s_1, 以此价格可以买入或卖出任意数量的股票。在这段时间内持有股票, 直到 $t=\delta t$, 期间股票的价格没有任何变化, 也没有任何费用支出, 期末股票将有一个新的价格。如果向下移动到节点 2, 价格将变为 s_2; 如果向上移动到节点 3 点, 价格将变为 s_3。

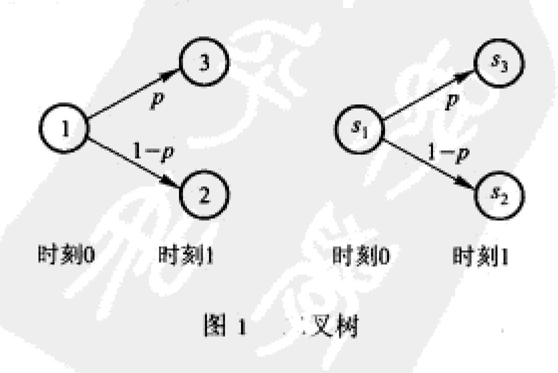


图 1　二叉树

2. 债券

现在考虑用债券价格描述货币的时间价值。从 $t=0$ 到 $t=\delta t$ 内, 设连续复利率为 r, 使得在时刻 0 时的 1 元到时刻 δt 时增加到 $\exp(r\delta t)$ 元, 而且以这个利率可以借入或贷出任意数量的货币。引入债券 B, 它在时刻 0 时的价格为 B_0, 而在 δt 后它的价格将变为 $B_0\exp(r\delta t)$。

以上两种金融工具就构成了一个金融市场。尽管它非常简单, 但是对于投资者来说, 仍存在某些不确定性。在金融工具市场中, 金融工具的价值依赖于 δt 时的股票价格。根据股票的未来价值, 金融工具的收益用函数 f 表示, 对应于节点 2 和节点 3 的奖励额或惩罚额分别为 $f(2)$ 和 $f(3)$。例如, 对履约价格为 k 的远期合约, $f(2)=s_2-k, f(3)=s_3-k$。

3. 无风险组合

现在的问题是: 什么类型的函数 f 能够通过一个合适的策略构造出来? 考虑由股票和债券构成的投资组合 (ϕ,ψ), 它包含 ϕ 数量的股票 S(价值为 ϕs_1)和 ψ 数量的债券(价值为 ψB_0)。如果在时刻 0 时买入这个投资组合, 那么成本是 $\phi s_1+\psi B_0$。

经过 δt 段时间后, 投资组合的价值将有两个可能值:

(i) 如果股票价格上升, 那么投资组合的价值为

$$\phi s_3+\psi B_0\exp(r\delta t)$$

(ii) 如果股票价格下降, 那么投资组合的价值为

$$\phi s_2+\psi B_0\exp(r\delta t)$$

为复制出 $f(3)$ 和 $f(2)$, 可调整 ϕ 和 ψ。因此要求组合策略满足下面两个方程

$$\phi s_3+\psi B_0\exp(r\delta t)=f(3)$$

$$\phi s_2+\psi B_0\exp(r\delta t)=f(2)$$

既然 S 是股票, $s_2\neq s_3$。对上述方程组求解, 得到

$$\phi=\frac{f(3)-f(2)}{s_3-s_2}$$
$$\psi=B_0^{-1}\exp(-r\delta t)\left(f(3)-\frac{(f(3)-f(2))s_3}{s_3-s_2}\right)$$

确定了 ϕ 和 ψ 后, 买入投资组合 (ϕ,ψ), 并且持有它, 那么上述两个方程表明: 如果股票价格上升, 那么投资组合的价值将变为 $f(3)$; 如果股票价格下降, 那么投资组合的价值将变为 $f(2)$。

在前面的简单模型中, 对任意衍生工具 f, 通过选择适当的债券和股票的投资组合, 可以预先构造出来。记 V 为买入投资组合 (ϕ,ψ) 的价格, 那么

$$\begin{aligned} V &= \phi s_1 + \psi B_0 \\ &= s_1\left(\frac{f(3)-f(2)}{s_3-s_2}\right) + \exp(-r\delta t)\left(f(3) - \frac{f(3)-f(2)}{s_3-s_2}\right) \end{aligned}$$

假设市场上衍生债券的交易价格为 P, 而且 $P<V$。如果投资者卖出投资组合 (ϕ,ψ), 得到 V, 同时以价格 P 买入衍生产品, 那么到时刻 δt 时, 衍生产品的价值将正好抵消投资组合的价值。这种同时卖出投资组合、买入衍生产品的收益为 $V-P$, 它是无风险套利。因此在市场上 P 将不是一个合理的价格, 市场会迅速地反应, 消除套利机会。

同样地, 假设 $P>V$。如果投资者以价格 P 卖出衍生产品, 同时买入投资组合 (ϕ,ψ), 成本为 V, 那么到时刻 δt 时, 衍生产品的价值将正好抵消投资组合的价值。这种同时卖出衍生产品、买入投资组合的收益为 $P-V$, 它是无风险套利。

总之, 只有 V 才是在时刻 0 时的唯一合理的价格。

4. 衍生工具价格的期望形式

上述 V 的表达式可以进一步写成如下形式:

$$V = \exp(-r\delta t)(qf(3) + (1-q)f(2))$$

其中

$$q = \frac{s_1\exp(r\delta t) - s_2}{s_3 - s_2}$$

通常假设 $s_3>s_2$。如果 $q\leqslant 0$, 那么 $s_1\exp(r\delta t)\leqslant s_2<s_3$, 其中 $s_1\exp(r\delta t)$ 是在期初价值为 s_1 的债券 B 在期末的价值。因此, 如果在期初出售债券, 同时以获得的资金买入股票, 那么将获得无风险套利。因此, 在任何市场上, 对二叉树的股票过程 S, 应有 $q>0$。

同样地, 如果 $q\geqslant 1$, 那么 $s_2<s_3\leqslant s_1\exp(r\delta t)$。此时在期初卖出股票, 同时买入债券将获得无风险套利。

总之, 一个合理的市场结构将使得 $0<q<1$。此时 q 可视为概率值, 对应的概率测度记为 $\mathbb{Q}$。

§2　二叉树模型

上一节讨论了非常简化的二叉树。如果能够进一步扩展二叉树, 可以得到一个更优的模型。本节将建立树模型。

这里的金融市场仍包括两种金融工具 —— 债券 B 和股票 S。金融工具的交易数量没有限制, 而且没有交易成本、信用风险以及买入卖出差价。现在考虑包含多个长度为 δt 的多周期模型。

1. 股票

股票 S 的价格变化是随机的, 但是这种随机性有一定的结构。在简化的二叉树模型中, 股票在期末时有两个可能值, 继续保持这种结构, 把各个分支连接起来构成一个树。多分支的树起始于一个简单的二叉树, 它出现在最初的第一个周期, 即从 $t=0$ 到 $t=\delta t$。如果在时刻 0 时 S 的价值是 $S_0=s_1$, 那么一个周期后的价值 S_1 有两个可能值: s_2 和 s_3。

股票在时刻 $t=\delta t$ 的价值确定后, 把二叉树结构扩展开来, 在第二个周期后, 股票价值又将有两个可能值。因此, 在第二个周期末, 股票价值有四个可能值。如从 s_2 开始, S_2 的可能值为 s_4 或 s_5; 如从 s_3 出发, S_2 的可能值为 s_6 或 s_7。

如图 2 所示, 在时刻 i, 股票有 2^i 个可能值。从节点 j 出发, 分支节点可能会下降到节点 $2j$ 或上升到节点 $2j+1$。

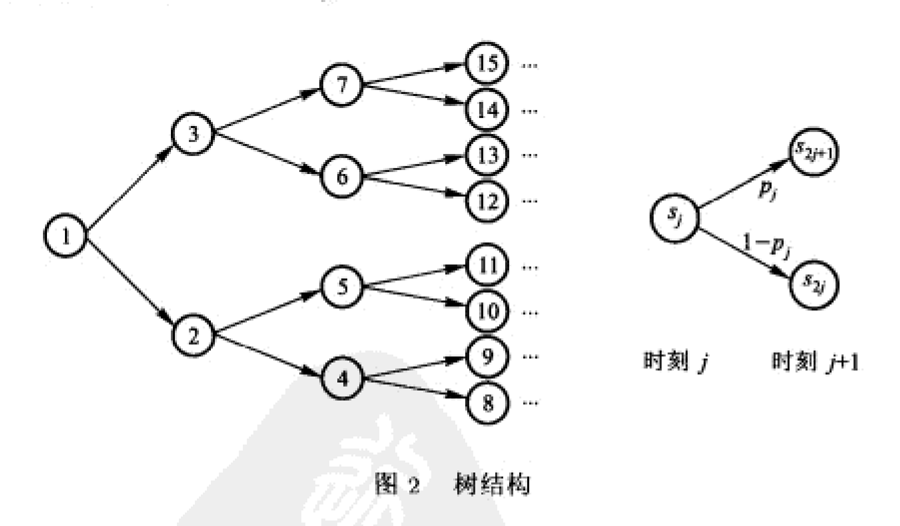


图 2 树结构

上述树结构有很大的灵活性。在实际问题中, 股票价格的可能值很多。对于某种股票, 如果认为在一个合适水平上描述股票价格的复杂性, 如假设股票价格在时刻 t 有 1000 个可能值, 那么仅需取 δt 足够小, 在时刻 t 之前得到包含 10 层或更多层的树。另外, 上述树结构有更广的概率结构。每一次向上或向下的选择都与相关的概率相联系。对每个二分支结构, 只需给出向上的概率 p_j 即可。给定初始值为 s_j, p_j 就是股票价格到达 s_{2j+1} 的概率, $1-p_j$ 就是股票价格到达 s_{2j} 的概率。

2. 债券

在简单的二叉树模型中, 能够完全预知债券的行为, 设利率为 r, 债券的价格增

长因子为 $\exp(r\delta t)$。实际上, 在整个树模型中, 利率不必是常数, 一般地可考虑利率序列, $R_0, R_1, \cdots$, 每个 R_i 在对应的周期初始是已知的。债券在时刻 $n\delta t$ 的价值就是 $B_0 \exp\left(\sum_{i=0}^{n-1} R_i \delta t\right)$。

因此, 对一般的债券, 它的价值也有一定的随机性, 但是与股票价格相比, 这种随机性是另一种不同类型的随机性。债券 B 和货币的时间价值有相同的结构。债券的利息在一段时间内会发生变化, 但是对一个周期而言, 债券的利息在期初是已知的, 这是因为利率在期初就已经确定。

为了简单, 在树模型中假设利率为常数 r, 此时, 在时刻 $n\delta t$ 时债券的价格就是 $B_0 \exp(r\delta t)$。

3. 向后推导方法

把上一节的结果扩展到整个二叉树模型中是很简单的。关键的思路就是向后推导方法 —— 把投资组合向后推导, 直到初始时刻。

例如, 考虑关于股票 S 的一般请求权。在简单的二叉树模型中, 请求权可表示为一个函数 f, 它的取值仅依赖于一个周期后的节点。这里扩展请求权的概念, 使得它不仅包括执行请求权时 S 的价值, 而且包括 S 的历史。

股票的树结构体现了节点与股票价格到该节点的历史路径的一一对应关系, 它把请求权价值与树的特定节点联系起来。

4. 二周期的树结构

图 3 给出了两阶段包含三个二分支结构的树模型, 这是最简单的树模型。

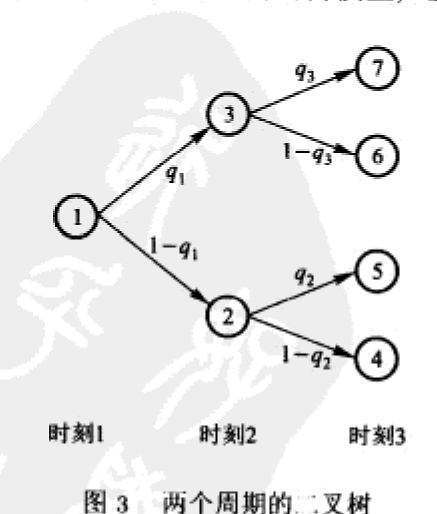


图 3 两个周期的二叉树

假设在每个周期内, 利率都是常数 r, 那么存在 q_j, 使得衍生工具在节点 j 的

价值 $f(j)$ 可表示为

$$f(j) = \mathrm{e}^{-r\delta t}(q_j f(2j+1) + (1-q_j)f(2j))$$

在图 3 中,

$$f(3) = \mathrm{e}^{-r\delta t}(q_3 f(7) + (1-q_3)f(6))$$

$$f(2) = \mathrm{e}^{-r\delta t}(q_2 f(5) + (1-q_2)f(4))$$

其中 $q_j = (s_j \exp(r\delta t) - s_{2j})/(s_{2j+1} - s_{2j})$, 具体地,

$$q_2 = \frac{s_2 \exp(r\delta t) - s_4}{s_5 - s_4}, q_3 = \frac{s_3 \exp(r\delta t) - s_6}{s_7 - s_6}$$

因此, 对第一个二分支结构来说, 如果股票价格上升, 那么在时刻 1 时的请求权价值为 $f(3)$; 如果股票价格下降, 那么在时刻 1 时的请求权价值为 $f(2)$。从而在时刻 0 的价值为

$$f(1) = \mathrm{e}^{-r\delta t}(q_1 f(3) + (1-q_1)f(2))$$

综合以上各个式子, 可以得到在时刻 0 时请求权的价值如下

$$f(1) = \mathrm{e}^{-2r\delta t}(q_1 q_3 f(7) + q_1(1-q_3)f(6) + q_2(1-q_1)f(5) + (1-q_1)(1-q_2)f(4))$$

5. 路径概率

树上的一条特定路径的概率是该路径所经过的各个分支概率的乘积。例如在图 3 中, 上升两次的概率是 $q_1 q_3$, 先上升后下降的概率为 $q_1(1-q_3)$, 这些路径概率对应的概率测度记为 $\mathbb{Q}$。

在两个周期的树结构中, 请求权的价格可表示为请求权变量关于概率测度 $\mathbb{Q}$ 的期望值与贴现因子 $\mathrm{e}^{-2r\delta t}$ 的乘积。

上述向后推导方法与结论可应用于一般的包含 n 个周期的树结构, 并得到类似的结论。下面给出一个例子。

例 1 图 4 称为重组树模型, 这是因为不同的分支可以在同一节点上会合。对这样的树, 计算比较简单。树上各节点的数字是股票价格 s, 在每个周期内股票价格上升的概率为 3/4, 下降的概率为 1/4。为了简单, 假设利率 $r=0$。

在时刻 3 时, 以 100 元买入股票, 求该看涨期权的价格。容易得到在时刻 3 时, 请求权的数值从上到下依次为 60, 20, 0, 0。

现在计算概率 q。此时利率 $r=0$, 风险中性概率 q 为

$$q = \frac{s_{\text{now}} - s_{\text{down}}}{s_{\text{up}} - s_{\text{down}}}$$

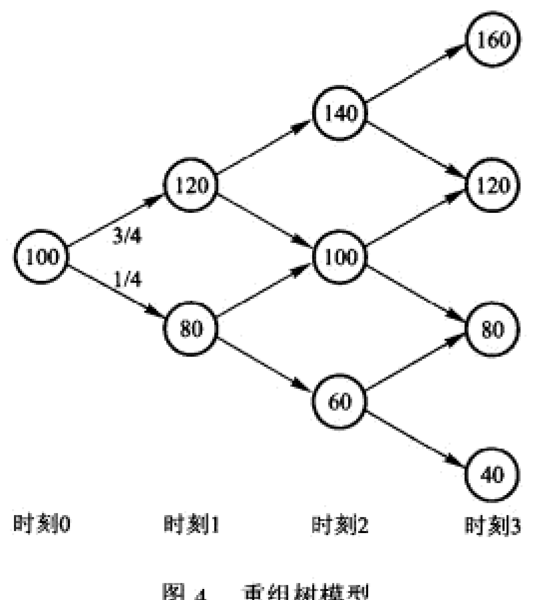


图 4　重组树模型

容易验证在每一个节点上, $q = 1/2$。

请求权的价值 f 为

$$f_{\text{now}} = q f_{\text{up}} + (1-q) f_{\text{down}}$$

把上述公式应用到在时刻 3 时相邻的两个节点上, 就可以得出在时刻 2 时的期权价值, 如图 5 所示。

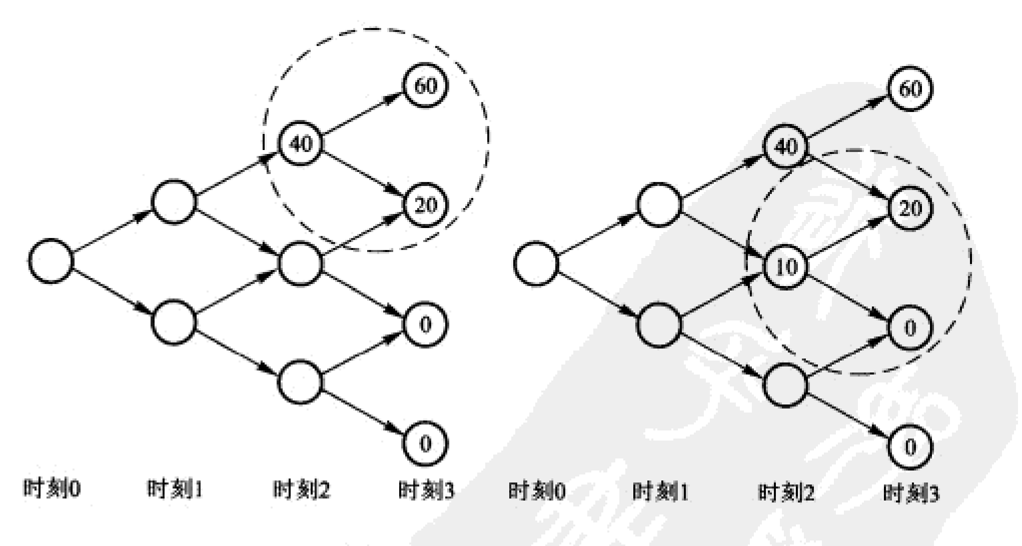


图 5　在时刻 2 时的期权价值

重复以上程序, 最后得到完整的树, 如图 6 所示。

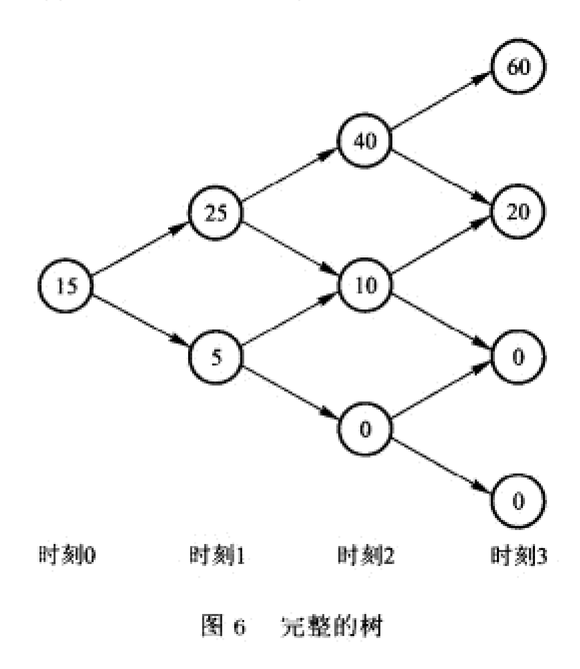


图 6 完整的树

在时刻 0 时期权的价格是 15。为考察套期保值, 在任何节点应用公式

$$\phi = \frac{f_{\text{up}} - f_{\text{down}}}{s_{\text{up}} - s_{\text{down}}}$$

ϕ 表示持有的股票单位数。

(1) 时刻 0: 期权价格是 15。计算 $\phi = (25-5)/(120-80) = 0.5$。买入 0.5 单位的股票, 成本为 50 元, 因此需要借入资金 35 元。

(2) 时刻 1: 假设股票价格上涨到 120 元。此时 $\phi = (40-10)/(140-100) = 0.75$, 因此以新的价格再买入 0.25 单位的股票, 额外借入资金 30 元, 总额达到 65 元。

(3) 时刻 2: 假设股票价格上涨到 140 元。此时 $\phi = (60-20)/(160-120) = 1$, 因此以新的价格再买入 0.25 单位的股票, 额外借入资金 35 元, 总额达到 100。

(4) 时刻 3: 假设股票价格下降到 120 元。现在用 1 单位的股票换取 120 元现金, 其中 100 元用来偿还债务, 剩余 20 元就是期权收益。

下面的表 1 显示了上述过程在一段时期内的变化。表中所描述的组合策略相应于各自的周期。例如, ϕ_1 是从 $i=0$ 到 $i=1$ 周期内持有股票的单位数。期权的价值与新旧两个投资组合的价值相同, 例如, $V_1 = \phi_1 S_1 + \psi_1 = \phi_2 S_1 + \psi_2$。

应用同样的分析, 如果股票价格在开始周期内下降到 80 元, 那么

(1) 时刻 1: $\phi = (10-0)/(100-60) = 0.25$, 此时卖出 0.25 个单位的股票, 得到现金 20 元, 债务降到 15 元。

(2) 时刻 2: 假设股票价格上升到 100 元, 那么 $\phi = (20-0)/(120-80) = 0.50$。此时买入 0.25 个单位的股票, 成本为 25 元, 从而债务增加到 40。

表 1　期权与投资组合的进展

时刻 i	上一次跳跃	股票价格 S_i	期权价值 V_i	持有股票单位 ϕ_i	债券持有量 ψ_i
0	—	100	15	—	—
1	上	120	25	0.50	-35
2	上	140	40	0.75	-65
3	下	120	20	1.00	-100

(3) 时刻 3: 假设股票价格下降到 80 元, 此时股票价值为 40 元, 正好抵消债务, 期权没有价值。

下面的表 2 显示了上述过程在一段时期内的变化。

表 2　在另一条不同的路径上投资组合的进展

时刻 i	上一次跳跃	股票价格 S_i	期权价值 V_i	股票持有 ϕ_i	债券持有 ψ_i
0	—	100	15	—	—
1	下	80	5	0.50	-35
2	上	100	10	0.25	-15
3	下	80	0	0.50	-40

注意到以上所有过程 (S, V, ϕ, ψ) 都依赖股票价格上下变动的路径。特别地, ϕ 和 ψ 都是随机的, 但是它们仅依赖于直到需要计算它们的时刻之前的路径。

在概率 $\mathbb{Q}$ 下, 对应于时刻 3 的每一节点的概率, 从上到下分别为: 1/8, 3/8, 3/8, 1/8。在这些概率下, 请求权的期望为 15 元。

6. 结论

树结构保证了在每个节点上任何请求权的价值仅有一个值, 否则套利就会出现。通过向后推导, 直到填满整个树。现在把本节涉及的内容与符号总结如下。

$$q = \frac{e^{r\delta t}s_{\text{now}} - s_{\text{down}}}{s_{\text{up}} - s_{\text{down}}}$$

$$f_{\text{now}} = e^{-r\delta t}(qf_{\text{up}} + (1-q)f_{\text{down}})$$

$$\phi = \frac{f_{\text{up}} - f_{\text{down}}}{s_{\text{up}} - s_{\text{down}}}$$

$$\psi = B_{\text{now}}^{-1}(f_{\text{now}} - \phi s_{\text{now}})$$

$$V = f(1) = \mathbb{E}_{\mathbb{Q}}(B_T^{-1}X)$$

这里

q : 向上运动的中性概率	r : 周期内的利率
f : 请求权价值过程	s : 股票价格过程
ϕ : 持有股票的策略	B : 债券价格过程, $B_0 = 1$
ψ : 持有债券的策略	$\mathbb{Q}$: 对应于q的测度
V : 在时刻0的请求权价值	X : 请求权的收益
δt : 时间周期长度	T : 请求权的期限

§3 二叉树表示定理

本节首先引入一些正式的定义。以下涉及七个不同的定义, 每个定义都将由包含 7 个节点的二叉树加以说明, 见图 7。

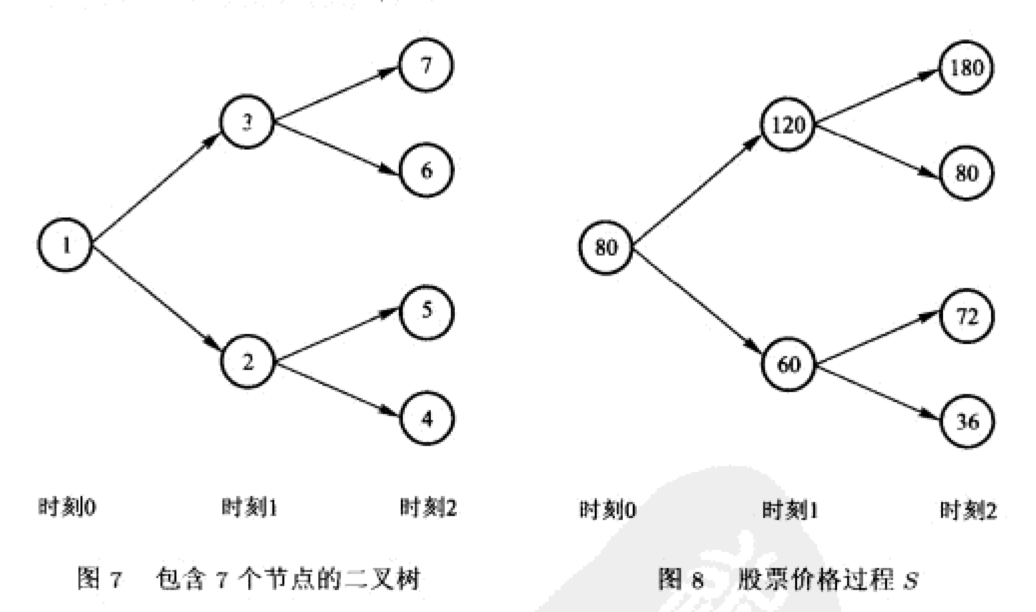


图 7 包含 7 个节点的二叉树　　　图 8 股票价格过程 S

1. 股票价格过程

在树的每个节点的股票价格的可能值, 按照时间顺序构成一个过程 S。图 8 描述了一个可能的过程。随机变量 S_i 表示在时刻 i 的股票价值。例如, 取决于节点 2 或节点 3, S_1 或者是 60, 或者是 120。

2. 概率测度

定义于树上的概率测度 $\mathbb{P}$ 或 $\mathbb{Q}$, 它们分别对应于概率 $\{p_j\}$ 或 $\{q_j\}$。概率测度与过程 S 是可分离的, p_j 表示过程从节点 j 向上移动的概率。图 9 给出了概率测

度 $\mathbb{P}$ 和 $\mathbb{Q}$ 的例子。

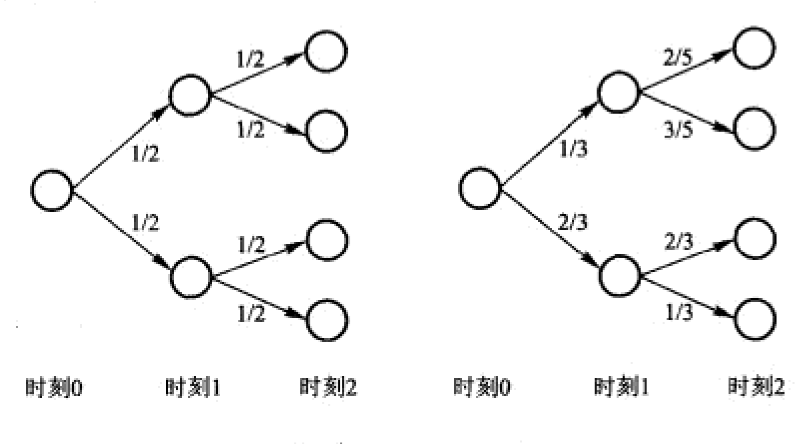


图 9　概率测度 $\mathbb{P}$ 和 $\mathbb{Q}$

3. 滤波

滤波 $(\mathcal{F}_i)$ 是股票价格直到时刻 i 的历史信息。在时刻 0 时的滤波 $\mathcal{F}_0$ 仅包括节点 1, 即 $\mathcal{F}_0 = \{1\}$。到时刻 1 时, 如果第一个周期内股票价格下降, $\mathcal{F}_1 = \{1,2\}$; 如果第一个周期内股票价格上升, $\mathcal{F}_1 = \{1,3\}$。对一般的 i, $\mathcal{F}_i$ 表示在时刻 i 时的节点的滤波。下表 3 给出了与每一节点相对应的滤波。

表 3　滤 波 过 程

节点	1	2	3	4	5	6	7
滤波	{1}	{1,2}	{1,3}	{1,2,4}	{1,2,5}	{1,3,6}	{1,3,7}

4. 请求权

请求权 X 是二叉树上最后时刻节点的函数。例如, 在时刻 2 的股票价值 S_2 是一个请求权。同样地, 对执行价格为 70 的看涨期权, 它在时刻 2 时的价值以及股票沿不同路径(对应于特定节点) 所达到的最高价格都是请求权。下表 4 给出了三种请求权变量的数值。

请求权和过程的本质区别在于: 请求权仅定义在最后时刻时的各个节点上, 过程定义于直到时刻 T 的各个节点上。

5. 条件期望

条件期望扩展了期望的概念, 它涉及两个参数, 即测度 $\mathbb{Q}$ 与历史 $\mathcal{F}_i$, 记为 $\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$。测度 $\mathbb{Q}$ 表明了如何决定路径概率, 从而决定期望。通常的期望是从时刻 0 出发沿着完整的路径来计算的, 在实际问题中, 有时需考虑从以后的时刻出发计

表 4 在时刻 2 的请求权数值

时刻 2的节点	S_2	$(S_2-70)^+$	$\max\{S_0,S_1,S_2\}$
7	180	110	180
6	80	10	120
5	72	2	80
4	36	0	80

算期望，为此滤波的概念就自然出现了。对于一个请求权 X, $\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$ 是 X 沿着时刻 i 后的路径的期望，初始路径片段为 $\mathcal{F}_i$。如果把过程在时刻 i 到达的节点视为树的新基点，那么条件期望就与通常的期望相一致。条件期望 $\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$ 依赖于滤波 $\mathcal{F}_i$ 的值，因此条件期望是一个随机变量。

对于时刻 i 的每个节点,$\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$ 可理解为在股票价格已经到达该节点的条件下 X 的期望。当 $X=S_2$, $\mathbb{P}$ 是由图 9 给出的测度时，下表 5 给出了 $\mathbb{E}_{\mathbb{P}}(S_2|\mathcal{F}_i)$。

表 5 关于滤波的条件期望

条件期望	滤波	数值
$\mathbb{E}_{\mathbb{P}}(S_2|\mathcal{F}_0)$	{1}	(180+80+72+36)/4=92
$\mathbb{E}_{\mathbb{P}}(S_2|\mathcal{F}_1)$	{1,3}	(180+80)/2=130
	{1,2}	(72+36)/2=54
$\mathbb{E}_{\mathbb{P}}(S_2|\mathcal{F}_2)$	{1,3,7}	180
	{1,3,6}	80
	{1,2,5}	72
	{1,2,4}	36

很自然地，关于初始节点的条件期望就是 $\mathbb{E}_{\mathbb{P}}(S_2)$。时刻 2 是所给定树的最后时刻，其后不再有新信息，所以 $\mathbb{E}_{\mathbb{P}}(S_2|\mathcal{F}_2)=S_2$ 成立。

$\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$ 构成了一个过程。当 $X=S_2$ 时,$\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_i)$ 由图 10 所示。按照这种方式，给定一个测度时，可以把请求权转化为过程。

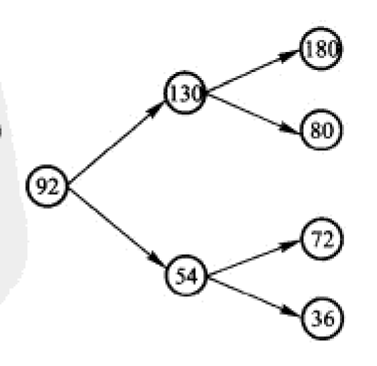


图 10 条件期望过程

6. 可预期过程

定义于树上的过程 $\phi=\{\phi_i\}$, 如果在时刻 i 时，变量 ϕ_i 的取值仅由 $\mathcal{F}_{i-1}$ 即可确定，那么称 ϕ 为可预期过程。可预期过程也是二叉树过程，在时刻 0 以后的每

个节点的价值都有定义。但是与基本的过程 S 相比, ϕ 的取值提前一个周期就已经确定。例如, 债券的价格过程 B_i 可以是可预期的。另外, 股票的延期价格过程 $\phi = S_{i-1}, i \geqslant 1$ 也是可预期的, 见图 11。

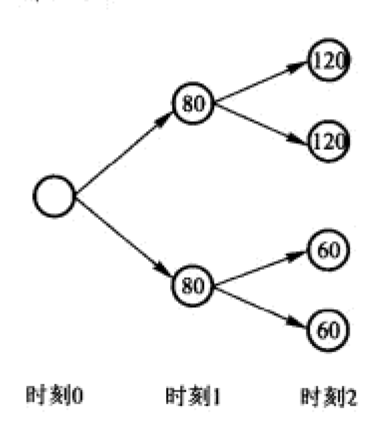


图 11 可预期过程

一般地, 交易策略都是可预期过程, 它的含义就是: 虽然不能事先知道股票价格, 但下一周期的交易策略通常由现在的信息来确定, 这是把套利机会(或内部人交易)排除在外的模型的一个本质特征。

在各种概念中, 以下关于鞅的定义可能是最重要的。在前两节中, 曾涉及与无风险投资组合相关的风险中性概率测度 $\mathbb{Q}$。一个很自然的问题就是: 如何描述测度 $\mathbb{Q}$? 为回答这个问题, 需要引入鞅的定义。

7. 鞅

定义 1 对于给定的测度 $\mathbb{P}$ 和滤波 $(\mathcal{F}_i)$, 如果过程 S 满足以下条件

$$\mathbb{E}_{\mathbb{P}}(S_j|\mathcal{F}_i) = S_i,\ i \leqslant j$$

那么称 S 为一个鞅。

按照上述定义, 在期望意义下, 过程 S 关于 $\mathbb{P}$ 没有向上和向下的偏差。如果在某节点过程的价值为 100, 那么在测度 $\mathbb{P}$ 下, 过程的条件期望也是 100。

例 2 取常数值的过程, 对于所有的概率测度而言, 都是平凡的鞅。

例 3 图 8 中的过程 S 关于图 9 中的测度 $\mathbb{Q}$ 是一个鞅。为验证此结论, 注意到

$$\mathbb{E}_{\mathbb{Q}}[S_1|\mathcal{F}_0] = \mathbb{E}_{\mathbb{Q}}[S_1] = (1/3) \times 120 + (2/3) \times 60 = 80 = S_0$$

如果在第一个周期内, 股票价格上升, 那么

$$\mathbb{E}_{\mathbb{Q}}[S_2|\mathcal{F}_1] = (2/3) \times 180 + (1/3) \times 80 = 120 = S_1$$

鞅定义的其他要求可类似验证。

例 4 条件期望过程 $N_i = \mathbb{E}_{\mathbb{P}}(S_2|\mathcal{F}_i)$ 是一个 $\mathbb{P}$- 鞅。由定义 1, 只需验证 $\mathbb{E}_{\mathbb{P}}[N_i|\mathcal{F}_0] = N_0$。实际上

$$\mathbb{E}_{\mathbb{P}}[N_i|\mathcal{F}_0] = (1/2) \times 130 + (1/2) \times 54 = 92 = N_0$$

例 4 是一般结论的具体例子, 即对于任何请求权 X, 过程 $\mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_i)$ 是 $\mathbb{P}$- 鞅。为证明此一般结论, 需要以下应用等式

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_j)|\mathcal{F}_i] = \mathbb{E}_{\mathbb{P}}[X|\mathcal{F}_i],\ i \leqslant j$$

上述等式就是关于条件期望的塔式法则。

根据塔式法则, 验证过程 S 是否为 $\mathbb{P}$- 鞅, 一个简单的方法就是比较 S_i 与终值 S_T 的条件期望 $\mathbb{E}_{\mathbb{P}}(S_T|\mathcal{F}_i)$ 是否相等。只有当 $\mathbb{E}_{\mathbb{P}}(S_T|\mathcal{F}_i) = S_i$ 时, S 才是 $\mathbb{P}$- 鞅。

鞅是依赖于概率测度的。同一个过程关于某个测度是鞅, 而关于另一个测度可以不是鞅。例如, 图 8 给出的过程 S 关于测度 $\mathbb{P}$ 不是鞅, 这只需注意到图 8 和图 10 是不同的即知。但 S 关于图 9 中的测度 $\mathbb{Q}$ 是鞅。测度 $\mathbb{Q}$ 称为 S 的鞅测度。

有了前面的准备后, 现在可以介绍二叉树表示定理, 该定理表述如下。

8. 二叉树表示定理

定理 假设二叉树价格过程 S 关于测度 $\mathbb{Q}$ 是鞅, N 是任意其他的 $\mathbb{Q}$- 鞅, 那么存在可预期过程 ϕ, 使得

$$N_i = N_0 + \sum_{k=1}^{i} \phi_k \Delta S_k$$

其中 $\Delta S_i = S_i - S_{i-1}$ 是 S 从时刻 $i-1$ 到 i 的变化量, ϕ_i 是在时刻 i 时 ϕ 的值。

上述定理表明, N_i 可从 N_0 出发, 可预期地得到。该定理可直接证明如下。

考虑一个从时刻 $i-1$ 到时刻 i 的二叉树。树的结构保证了 $\mathcal{F}_i$ 在 $\mathcal{F}_{i-1}$ 之后有两种选择, 分别相应于股票价格的上升和下降。过程 S 和 N 的增量分别为

$$\Delta S_i = S_i - S_{i-1}, \qquad \Delta N_i = N_i - N_{i-1}$$

这些增量所包含的波动性对应于二叉树本身的几何宽度, 如图 12 所示。

由于每个二叉树仅有两条路径, 因此依赖于二叉树的任何随机变量都由两个因素完全确定: 二叉树的宽度与仅依赖于 $\mathcal{F}_{i-1}$ 的常数偏移量。如果从一个随机过程出发, 构造另一个随机过程, 一般地将基于标度(与宽度相适应)和平移(与偏移相适应)两个方面。

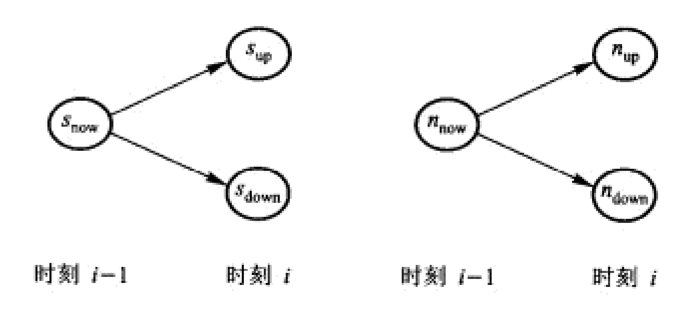


图 12　鞅的波动性与二叉树的几何宽度

首先考虑标度。对于 S 来说, 向上和向下的价格之差为 $\delta s_i = s_{\text{up}} - s_{\text{down}}$。同样地, 对于 N 来说, 价格之差为 $\delta n_i = n_{\text{up}} - n_{\text{down}}$。$\delta s_i$ 和 δn_i 都仅依赖于滤波 $\mathcal{F}_{i-1}$。定义 ϕ_i 为二叉树的宽度之比

$$\phi_i = \frac{\delta n_i}{\delta s_i}$$

其次考虑平移。增量 ΔN_i 与 ΔS_i 之间应有如下关系

$$\Delta N_i = \phi_i \Delta S_i + k$$

其中 ϕ_i 和平移常数 k 都由 $\mathcal{F}_{i-1}$ 决定。

另一方面, 由于 S 和 N 都是 $\mathbb{Q}$- 鞅, 因此 $\mathbb{E}_{\mathbb{Q}}[\Delta N_i|\mathcal{F}_{i-1}] = \mathbb{E}_{\mathbb{Q}}[\Delta S_i|\mathcal{F}_{i-1}] = 0$。标度因子 ϕ_i 是可预期的, 它在时刻 $i-1$ 时就已经是已知的, 所以 $\mathbb{E}_{\mathbb{Q}}[\phi_i \Delta S_i|\mathcal{F}_{i-1}] = 0$。在 ΔN_i 与 ΔS_i 之间关系式两边取条件期望, 可得 $k = 0$。

总之, 当 S 和 N 都是 $\mathbb{Q}$- 鞅时, $\Delta N_i = \phi_i \Delta S_i$。再考虑所有的节点后, 即得定理结论。

9. 二叉树表示定理在金融上的应用

在上述关于二叉树过程的定理中, 没有涉及股票和债券的投资组合, 也没有考虑套利机会及其对市场的影响。现在考虑如何把二叉树表示定理应用于定价问题。

在关于市场的二叉树模型中, 股票价格 S 构成二叉树过程。如果存在测度 $\mathbb{Q}$, 使得 S 为 $\mathbb{Q}$- 鞅, 那么可以应用二叉树表示定理, 用股票价格表示另外的鞅 N。定理中的可预期 ϕ 过程对应于构造策略, 它的含义是在每一个节点买入 ϕ_i 单位的股票。

如果逐步进行下去, 鞅 N 的终值正好等于某请求权, 那么请求权也就能够被复制出来。为此需要注意到以下两点。

首先, 请求权 X 只是一个随机变量, 它不是鞅。其次, 市场上既有股票, 也有债券。直观上, 二叉树表示定理中的 ϕ_i 将是构造策略的一个重要部分, 它表示股票的持有量。但同时也要考虑 ψ_i, 它表示债券的持有量。

现把请求权 X 变量转化成一个鞅。对给定的测度 $\mathbb{Q}$, 通过取条件期望, 可以构造过程

$$E_i = \mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_i]$$

因此, 如果存在 $\mathbb{Q}$, 使得 S 是 $\mathbb{Q}$- 鞅, 那么选择这样的 $\mathbb{Q}$ 后, E 也是一个 $\mathbb{Q}$- 鞅。

债券 B_i 表示货币价值的增长。在时刻 0 时的 1 元等价于在时刻 i 时的 B_i 元。B_i 是可预期的过程。不失一般性, 假设 $B_0 = 1$。

(i) 过程 B_i^{-1} 是另一个可预期的过程, 称为贴现过程。

(ii) 定义 $Z_i = B_i^{-1}S_i$, 称为贴现股票价格过程。

(iii) $B_T^{-1}X$ 也是一个请求权, 称为贴现请求权。

10. 构造策略

假设存在测度 $\mathbb{Q}$, 使得 Z 是 $\mathbb{Q}$- 鞅。设 X 为请求权, 那么 $E_i = \mathbb{E}_{\mathbb{Q}}[B_T^{-1}X|\mathcal{F}_i]$ 也是 $\mathbb{Q}$- 鞅。由二叉树表示定理, 存在一个可预期过程 ϕ, 使得

$$E_i = E_0 + \sum_{k=1}^{i} \phi_k \Delta Z_k$$

现在考虑以下构造策略：在时刻 i, 买入投资组合 Π_i, 它的组成如下：

(i) ϕ_{i+1} 单位的股票。

(ii) $\psi_{i+1} = (E_i - \phi_{i+1}B_i^{-1}S_i)$ 单位的债券。

在时刻 0, Π_0 的价值为 $\phi_1 S_0 + \psi_1 B_0$, 代入 ϕ_1 和 ψ_1 的表达式, 即得 $\phi_1 S_0 + \psi_1 B_0 = E_0$, 由 E 的定义, 即得 $E_0 = \mathbb{E}_{\mathbb{Q}}[B_T^{-1}X]$, 因此初始投资组合的成本为 $\mathbb{E}_{\mathbb{Q}}[B_T^{-1}X]$。因为 ϕ 和 ψ 都是可预期的, 所以容易确定 ϕ_1 和 ψ_1 的值。

经过一个周期后, 投资组合的价值发生了变化, Π_0 的值变为

$$\begin{aligned}
\phi_1 S_1 + \psi_1 B_1 &= B_1(E_0 + \phi_1(B_1^{-1}S_1 - B_0^{-1}S_0)) \\
&= B_1(E_0 + \phi_1 \Delta Z_1) \\
&= B_1 E_1
\end{aligned}$$

上面的最后一式用到了二叉树表示定理。在时刻 1, Π_0 的值变为 B_1E_1。

在时刻 1, 构造策略要求买入一个新的投资组合 Π_1, 但它的成本正是 B_1E_1。因此可以从投资组合 Π_0 得到 Π_1。依次类推, 在时刻 i, 投资组合 Π_i 的买入成本为 B_iE_i, 它在时刻 $i+1$ 的价值变为 $B_{i+1}E_{i+1}$。最后, 到时刻 T 时, 最后一个投资组合 Π_{T-1} 的价值变为 $B_TB_T^{-1}X = X$, 这正是请求权 X。

上面的构造策略称为“自融资的”, 它的含义稍后解释。

11. 无套利性

由前面的构造策略, 可知请求权 X 在时刻 0 的价格就是 $E_0 = \mathbb{E}_{\mathbb{Q}}[B_T^{-1}X]$, 即价格可表示为贴现请求权关于测度 $\mathbb{Q}$ 的期望, 其中测度 $\mathbb{Q}$ 是贴现股票过程 Z 的鞅测度。E_0 是无套利价格, 这是因为对任何其他价格, 通过构造 (ϕ_i, ψ_i) 策略, 可复制请求权, 进而会导致套利机会。

12. 自融资策略的存在性

在二叉树模型中, 对任意请求权, 总可以构造复制该请求权的自融资策略 (ϕ_i, ψ_i)。为解释自融资的含义, 以 V_i 表示交易策略在时刻 i 时投资组合 Π_i 的开始价值, 即 $V_i = \phi_{i+1}S_i + \psi_{i+1}B_i$。如果 Π_{i-1} 在时刻 i 的终止价值 $\phi_i S_i + \psi_i B_i$ 与 V_i 相等, 那么交易策略就是自融资的。换言之, 如果以 D_i 表示构造投资组合 Π_i 时需要补充的资金, 那么 $D_i = 0$, 这里

$$D_i = V_i - \phi_i S_i - \psi_i B_i$$

另一种表示自融资的方式可通过策略价值过程的变化来表示。

$$\Delta V_i = V_i - V_{i-1} = \phi_i \Delta S_i + \psi_i \Delta B_i + D_i$$

在时刻 i 时, $D_i = 0$ 等价于策略价值从时刻 $i-1$ 到 i 的变化仅来自于股票和债券价值的变化。

下面给出自融资策略的严格定义。

定义 2　给定关于市场的二叉树模型, 它包括股票 S 和债券 B, 如果 (ϕ_i, ψ_i) 满足以下三个条件, 那么就称 (ϕ_i, ψ_i) 是构造请求权 X 的自融资策略。

(i) ϕ 和 ψ 都是可预期过程。

(ii) 投资组合的价值变化满足方程 $\Delta V_i = \phi_i \Delta S_i + \psi_i \Delta B_i$, 其中 $\Delta S_i = S_i - S_{i-1}$, $\Delta B_i = B_i - B_{i-1}$。

(iii) $\phi_T S_T + \psi_T B_T = X$。

13. 在鞅测度下求贴现请求权的期望

在二叉树模型中, 任何衍生产品的价格是贴现请求权关于测度 $\mathbb{Q}$ 的期望, 其中测度 $\mathbb{Q}$ 是贴现股票价格过程的鞅测度。一般地, 期限为 T 的请求权 X 在时刻 i 的价值为

$$B_i E_i = B_i \mathbb{E}_{\mathbb{Q}}[B_T^{-1} X | \mathcal{F}_i]$$

14. 测度$\mathbb{Q}$ 的存在性和唯一性

最后指出, 在离散模型下, 对于任何理性的股票过程 S, 存在唯一的测度 $\mathbb{Q}$, 使得贴现股票价格过程 $B_i^{-1}S_i$ 是 $\mathbb{Q}$- 鞅。直观上讲, 在每一个二叉树上, 存在概率 $(q_i, 1-q_i)$, 把样本路径上对应的概率相乘, 即可构造出测度 $\mathbb{Q}$。

为了理解概率测度 $\mathbb{P}$ 和 $\mathbb{Q}$ 的区别, 现在考虑 $t=1$ 股票价格的期望值。在概率测度 $\mathbb{P}$ 下,

$$\mathbb{E}_{\mathbb{P}}[S_1] = S_0 \times (pu + (1-p)d)$$

而在概率测度 $\mathbb{Q}$ 下,

$$\mathbb{E}_{\mathbb{Q}}[S_1] = S_0 \times (qu + (1-q)d) = S_0 \times \left(\frac{u(\mathrm{e}^r - d)}{u-d} + \frac{d(u-\mathrm{e}^r)}{u-d}\right) = S_0\mathrm{e}^r$$

在概率测度 $\mathbb{Q}$ 下, 有风险的股票的期望收益与无风险投资的收益相同。换言之, 在概率测度 $\mathbb{Q}$ 下, 投资者关于风险是中性的, 投资者不会因为承担更多的风险从而要求额外的收益。因此, 有时也称 $\mathbb{Q}$ 为风险中性概率测度。

在实际概率测度 $\mathbb{P}$ 下, 股票的期望收益与无风险投资的收益通常是不同的。投资于有风险的股票, 投资者就要求有更高的期望收益。

现在考虑有 n 个周期的衍生工具合约的定价问题。如果在到期日 n 时股票的价格为 x, 那么以 $f(x)$ 表示衍生工具的支付额。例如, 对于欧式看涨期权, $f(x) = \max\{x-K, 0\}$, 其中 K 为执行价格。

假定在每一周期, 股票的价格或者上升(上升因子为 u), 或者下降(下降因子为 $d=1/u$)。此外, 无风险利率取常数值 r, 而且 $d < \mathrm{e}^r < u$。首先,

$$S_t = S_0 u^{N_t} d^{t-N_t}$$

其中 N_t 是在时刻 0 到 t 之间股票价格上升的次数。由假设可知, 在时刻 n 时有 $n+1$ 种可能的状态。另外, 在时刻 t 时的股票价格仅依赖于价格上升与下降的次数, 而与上升和下降所发生的次序无关。由于以上性质, 上述模型也称为重组二叉树模型或二分支格模型, 如图 13 所示。

二叉树模型也称为随机游动模型, 它的样本空间 Ω 由从时刻 0 到 n 的所有样本路径组成。因为在每一周期股票价格有 2 种可能, 所以共有 2^n 条样本路径。信息 $\mathcal{F}$ 是由从时刻 0 到 n 的各种样本路径的集合组成的, 而滤波流 $\mathcal{F}_t$ 是由到时刻 t 时为止的各种样本路径的集合组成的。给定样本空间 Ω, 到时刻 t 时为止的每一条样本路径都对应于样本空间的 2^{n-t} 个元素, 其中每一个元素在时刻 0 到 t 时都是相同的。N_t, S_t 都是随机变量, 它们都是样本空间 Ω 上的函数。

在这个模型下, 在所有周期, 价格上升的概率 q 都是一样的, 而且各个周期的价格变化都是相互独立的。因此, 在概率测度 $\mathbb{Q}$ 下, 到时刻 t 时为止价格上升的次

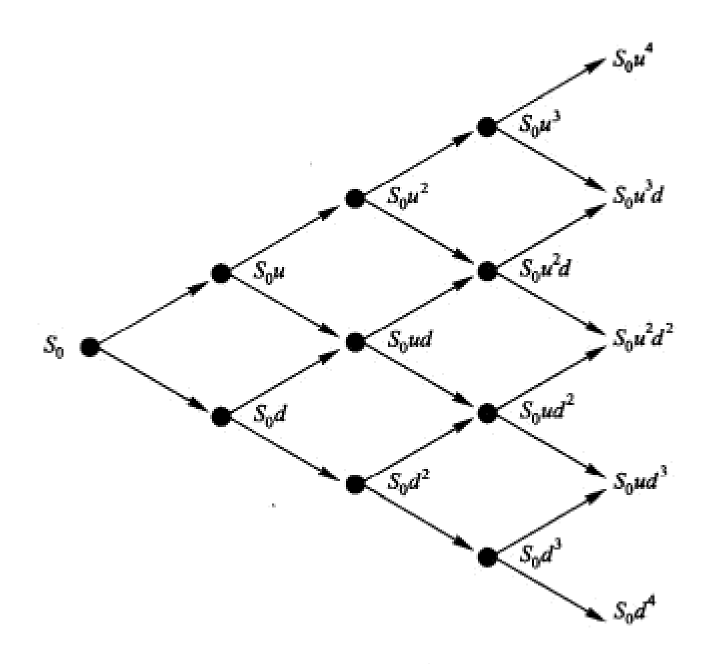


图 13　重组二叉树模型

数服从参数为 (t,q) 的二项分布。此外, 对于 $0<t<n$, N_t 与 N_n-N_t 独立, 而且 N_n-N_t 服从参数为 $(n-t,q)$ 的二项分布。

进一步, 引入另外一些符号。在给定 $N_t=j$ 的条件下, 其中 $j=0,1,\cdots,t$, $V_t(j)$ 表示衍生工具在时刻 t 时处于状态 j 下的价值, $V_t=V_t(N_t)$ 表示衍生工具在时刻 t 时的价值, 它是一个随机变量。因此, $V_n(j)=f(S_0u^jd^{n-j})$。

为了计算在时刻 0 时的价值 $V_0(0)$, 先从时刻 n 开始, 每次考虑一个周期, 应用前面的单周期二叉树模型, 逆向计算。

首先考虑从 $n-1$ 到 n 的最后一个周期。假设 $N_{n-1}=j$, 类似于单周期模型,

$$\begin{aligned}V_{n-1}(j)&=\mathrm{e}^{-r}[qV_n(j+1)+(1-q)V_n(j)]\\&=\mathrm{e}^{-r}\mathbb{E}_{\mathrm{Q}}[V_n|N_{n-1}=j]\\&=\mathrm{e}^{-r}\mathbb{E}_{\mathrm{Q}}[f(S_n)|\mathcal{F}_{n-1},N_{n-1}=j]\end{aligned}$$

其中

$$q=\frac{\mathrm{e}^r-d}{u-d}$$

而且上式的推导中用到了 Markov 性。上式可改写为

$$V_{n-1}=\mathrm{e}^{-r}\mathbb{E}_{\mathrm{Q}}[f(S_n)|\mathcal{F}_{n-1}]$$

进一步逆向推导, 并应用条件期望的塔式法则, 可得

$$\begin{aligned}
V_{t-1} &= \mathrm{e}^{-r}\mathbb{E}_{\mathbb{Q}}[V_t|\mathcal{F}_{t-1}] \\
&= \mathrm{e}^{-r}\mathbb{E}_{\mathbb{Q}}[\mathrm{e}^{-r}\mathbb{E}_{\mathbb{Q}}(V_{t+1}|\mathcal{F}_t)|\mathcal{F}_{t-1}] \\
&= \mathrm{e}^{-2r}\mathbb{E}_{\mathbb{Q}}[V_{t+1}|\mathcal{F}_{t-1}] \\
&\vdots \\
&= \mathrm{e}^{-(n-t+1)r}\mathbb{E}_{\mathbb{Q}}[V_n|\mathcal{F}_{t-1}] \\
&= \mathrm{e}^{-(n-t+1)r}\mathbb{E}_{\mathbb{Q}}[f(S_n)|\mathcal{F}_{t-1}]
\end{aligned}$$

最后得到

$$V_0 = \mathrm{e}^{-nr}\mathbb{E}_{\mathbb{Q}}[f(S_n)|\mathcal{F}_0] = \mathrm{e}^{-nr}\mathbb{E}_{\mathbb{Q}}[f(S_n)|S_0]$$

衍生工具在时刻 t 时的价格为

$$\begin{aligned}
V_t &= \mathrm{e}^{-r(n-t)}\mathbb{E}_{\mathbb{Q}}[f(S_t u^{N_n-N_t} d^{(n-t)-(N_n-N_t)})|N_t] \\
&= \mathrm{e}^{-r(n-t)}\sum_{k=0}^{n-t} f(S_t u^k d^{(n-t)-k})\frac{(n-t)!}{k!(n-t-k)!}q^k(1-q)^{n-t-k}
\end{aligned}$$

在单周期模型中, 由于 $\mathbb{E}_{\mathbb{Q}}[S_1] = S_0\mathrm{e}^r$, 从而称 $\mathbb{Q}$ 为风险中性概率测度。同样地, 对于多周期模型, 反复进行逆向推导, 可得

$$\mathbb{E}_{\mathbb{Q}}[S_t] = S_0\mathrm{e}^{rt}$$

所以风险中性概率测度的表述在多周期模型中仍然有效。另一方面, 由上面的推导方法可得

$$\mathbb{E}_{\mathbb{Q}}[\mathrm{e}^{-rT}S_T|\mathcal{F}_t] = \mathrm{e}^{-rt}S_t$$

上式表明贴现的资产价值过程 $D_t = \mathrm{e}^{-rt}S_t$ 关于概率测度 $\mathbb{Q}$ 是一个鞅, 因此又称 $\mathbb{Q}$ 为等价鞅测度。

第十一章　连续时间金融模型

§1　鞅表示定理

对连续过程, 应用 Itô 公式可以求解某些随机微分方程, 而当概率测度变化时随机微分方程的形式也发生变化。在离散过程中解决定价问题的关键是鞅测度的概念, 衍生工具的价格是收益变量关于鞅测度的期望, 而且由鞅测度可得交易策略, 从而验证了衍生工具的价格。在连续过程中, 也有类似的结论。为此先给出鞅的概念。

1. 鞅的概念

给定随机过程 $\{M_t, t \geqslant 0\}$ 与概率测度 $\mathbb{P}$, 如果以下条件成立, 那么就称 $\{M_t, t \geqslant 0\}$ 为 $\mathbb{P}$- 鞅。

(i) 对所有的 t, $\mathbb{E}_{\mathbb{P}}(|M_t|) < \infty$。

(ii) 对所有的 $s \leqslant t$, $\mathbb{E}_{\mathbb{P}}(M_t|\mathcal{F}_s) = M_s$。

按照鞅的定义, 在现在和过去历史条件下, 未来变量的条件期望等于现在的取值。

例 1　常数过程 $S_t \equiv c$ 关于任意概率测度都是鞅。

例 2　$\mathbb{P}$-Brown 运动是 $\mathbb{P}$- 鞅。这是因为 Brown 运动的增量 $W_t - W_s$ 与 $\mathcal{F}_s$ 独立, 而且 $W_t - W_s \sim N(0, t-s)$, 从而$\mathbb{E}_{\mathbb{P}}(W_t - W_s|\mathcal{F}_s) = 0$,

$$\mathbb{E}_{\mathbb{P}}(W_t|\mathcal{F}_s) = \mathbb{E}_{\mathbb{P}}(W_s|\mathcal{F}_s) + \mathbb{E}_{\mathbb{P}}(W_t - W_s|\mathcal{F}_s) = W_s$$

例 3 对于任意的仅依赖于直到时刻 T 的事件的请求权 X, 假设 $\mathbb{E}_{\mathbb{P}}(|X|) < \infty$, 那么过程 $N_t = \mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_t)$ 是一个 $\mathbb{P}$- 鞅。

首先 $N_t = \mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_t)$ 定义了一个过程, 为使 N_t 是 $\mathbb{P}$- 鞅, 需要 $\mathbb{E}_{\mathbb{P}}(N_t|\mathcal{F}_s) = N_s$ 成立, 它就是如下表达式

$$\mathbb{E}_{\mathbb{P}}(\mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_t)|\mathcal{F}_s) = \mathbb{E}_{\mathbb{P}}(X|\mathcal{F}_s)$$

上式是关于条件期望的塔式法则。

2. 鞅表示定理

对连续过程, 同样有鞅表示定理, 表述如下。

定理 假设 $\{M_t, 0 \leqslant t \leqslant T\}$ 是一个 $\mathbb{Q}$- 鞅, 它有如下形式

$$\mathrm{d}M_t = \sigma_t \mathrm{d}W_t + \mu_t \mathrm{d}t$$

σ_t 称为波动参数, $\sigma_t \neq 0$。μ_t 称为漂移系数。如果 N_t 是另外的一个 $\mathbb{Q}$- 鞅, 那么存在一个 $\mathcal{F}$- 可预期过程 ϕ, 使得 $\int_0^T \phi_t^2 \sigma_t^2 \mathrm{d}t < \infty$, 而且$N$有如下表示

$$N_t = N_0 + \int_0^t \phi_s \mathrm{d}M_s$$

进一步, ϕ 是唯一的。过程 ϕ 是 N 和 M 相应的波动参数的比值。

3. 无漂移项

一般的随机微分方程中包含了漂移项。一个很自然的问题是: 鞅的微分是否总能表示成 $\sigma_t \mathrm{d}W_t$ 的形式? 其中 σ_t 是某个 $\mathcal{F}$- 可预期过程。反过来, 没有漂移项的随机过程是否总是鞅?

应用鞅表示定理, 如果过程 X_t 是 $\mathbb{P}$- 鞅, 那么对 $\mathbb{P}$-Brown 运动 W_t, 存在 $\mathcal{F}$- 可预期过程 ϕ_t, 使得

$$X_t = X_0 + \int_0^t \phi_s \mathrm{d}W_s$$

上式是增量 $\mathrm{d}X_t = \phi_t \mathrm{d}W_t$ 的积分形式, 没有漂移项。

反过来, 如果 X 的微分形式为 $\mathrm{d}X_t = \sigma_t \mathrm{d}W_t + \mu_t \mathrm{d}t$, 而且

$$\mathbb{E}\left[\left(\int_0^T \sigma_s^2 \mathrm{d}s\right)^{1/2}\right] < \infty$$

那么 X 是鞅等价于漂移项为 0, 即 $\mu_t \equiv 0$。

4. 指数鞅

对上述结论中的条件, 有时候验证起来并不方便。例如, 设无漂移项的随机微分方程为 $\mathrm{d}X_t = \sigma_t X_t \mathrm{d}W_t$, 此时很难验证 $\mathbb{E}\left[\left(\int_0^T \sigma_s^2 X_s^2 \mathrm{d}s\right)^{1/2}\right] < \infty$, 更实用的判断方法如下。

如果 $\mathrm{d}X_t = \sigma_t X_t \mathrm{d}W_t$, 其中 σ_t 是 $\mathcal{F}$- 可预期过程, 那么当下列条件成立时

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\int_0^T \sigma_s^2 \mathrm{d}s\right)\right) < \infty$$

X 是一个鞅。另外上述随机微分方程的解是

$$X_t = X_0 \exp\left(\int_0^t \sigma_s \mathrm{d}W_s - \frac{1}{2}\int_0^t \sigma_s^2 \mathrm{d}s\right)$$

§2 构造策略

至此已有了各种数学工具——Itô 公式、Girsanov 定理以及鞅表示定理, 下一步需要按照一定的思路, 把它们应用于金融模型中。在最简单的 Black-Scholes 模型中, 市场包括两种资产: 风险证券和无风险债券。

下面首先引入投资组合的概念。

1. 投资组合 (ϕ, ψ)

投资组合可用两个过程 ϕ 和 ψ 来描述, 它们分别表示在时刻 t 时持有的证券单位数与债券数量。过程可以取负值, 即允许卖空股票或债券。ϕ 应是 $\mathcal{F}$- 可预期过程, 它仅依赖于到时刻 t 前的信息, 但不包括在时刻 t 的信息。

如果 ϕ 是左连续的, 那么 ϕ 是可预期的。

2. 自融资策略

投资组合的整体构成投资策略。投资策略是动态的, 在时刻 t 时的投资组合记为 (ϕ_t, ψ_t)。特别有意义的投资策略是自融资策略。

投资组合是自融资的, 当且仅当它的价值变化只依赖于资产价格的变化。在离散模型下它是通过差分方程给出的, 而在连续模型下, 它等价于某个随机微分方程。

3. 随机微分方程

给定股票价格 S_t 和债券价格 B_t, 投资组合 (ϕ_t, ψ_t) 在时刻 t 时的价值 V_t 可表示为 $V_t = \phi_t S_t + \psi_t B_t$。在下一时刻, 一方面 S_t 和 B_t 的价值变化会导致旧的投

资组合的价值变化, 另一方面, 旧的投资组合要调整为一个新的投资组合。如果调整的成本正好由投资组合所产生的利润或损失来支持, 不需要额外的资本, 这样的投资组合就是自融资的。

在离散时间情况下, 自融资的条件可表示为差分方程

$$\Delta V_i = \phi_i \Delta S_i + \psi_i \Delta B_i$$

而在连续时间情况下, 自融资的条件可表示为如下随机微分方程

$$\mathrm{d}V_t = \phi_t \mathrm{d}S_t + \psi_t \mathrm{d}B_t$$

现在考虑两个具体的例子。假设股票价格过程 S 是标准 Brown 运动 W, 即$S = W$, 债券价格过程 B 是常数过程, $B_t \equiv 1$, 下面给出了两个自融资的投资策略。

例 4 假设对于所有的 t, $\phi_t = \psi_t = 1$。如果在任意时刻 t 总持有一个单位的股票和一个单位的债券, 那么投资组合的价值 $V_t = W_t + 1$, 此价格会发生变化, 但是这种变化完全是由股票价格的波动造成的。直观上, 保持 (ϕ, ψ) 策略时不需要额外资本流入或投资组合的资本流出, (ϕ, ψ) 应该是自融资的。

如果从形式上加以验证, 那么由 $V_t = W_t + 1$ 即得 $\mathrm{d}V_t = \mathrm{d}W_t$, 它就是 $\mathrm{d}V_t = \phi_t \mathrm{d}S_t + \psi_t \mathrm{d}B_t$, 此处 $\mathrm{d}B_t = 0$。

例 5 假设 $\phi_t = 2W_t$ 和 $\psi_t = -t - W_t^2$。(ϕ, ψ) 是一个投资组合, ϕ 是可预期的, 该组合的价值 $V_t = \phi_t S_t + \psi_t B_t = W_t^2 - t$。由 Itô 公式, $\mathrm{d}V_t = 2W_t \mathrm{d}W_t$, 它就是 $\mathrm{d}V_t = \phi_t \mathrm{d}S_t + \psi_t \mathrm{d}B_t$。

第二个例子顺便说明了自融资并不是任意一个投资组合自动就有的性质。

4. 可复制策略

假设市场中仅包含无风险债券 B 和有风险资产 S, S 的波动参数为 σ_t, 请求权 X 依赖于到时刻 T 时的事件。

关于 X 的可复制策略是指自融资策略 (ϕ, ψ), 使得

$$\int_0^T \sigma_t^2 \phi_t^2 \mathrm{d}t < \infty, \ V_T = \phi_T S_T + \psi_T B_T = X$$

如果存在一个可复制策略 (ϕ, ψ), 那么 X 在时刻 t 时的价格必是 $V_t = \phi_t S_t + \psi_t B_t$。否则, 假设 X 在时刻 t 时的价格 P_t 低于 V_t, 那么市场交易者可以在时刻 t 卖出 ϕ_t 个单位的 S 和 ψ_t 个单位的 B, 得到 V_t, 此后保持空头 (ϕ, ψ) 直到时刻 T。同时, 在时刻 t 时以价格 P_t 买入请求权。这样在时刻 t 时, 市场交易者的收益为 $V_t - P_t$。因为 (ϕ, ψ) 是可复制策略, 它在时刻 T 时的价值是 X。到时刻 T 时, 买入的请求权和卖出的投资组合将会抵消, 而且在时刻 t 和 T 之间不需要额外的资本。在时刻 t 时产生的利润是无风险的, 这是套利机会。

同样地, 如果 X 在时刻 t 时的价格 P_t 高于 V_t, 也可以构造出套利机会。

在市场模型中出现的股票价格过程可以较为复杂。对于请求权 X, 需寻找可复制策略, 从而确定请求权的价格。

§3 Black-Scholes 模型

本节对一个具体的股票模型, 首先应用 Girsanov 定理, 把股票过程转化为一个新测度下的鞅, 然后应用鞅表示定理, 对每个请求权构造可复制策略。

1. 基本的 Black-Scholes 模型

假设存在常数 r, μ, σ, 使得股票价格 S_t 和债券价格 B_t 满足下面的关系

$$S_t = S_0 \exp(\sigma W_t + \mu t)$$
$$B_t = \exp(rt)$$

其中 r 是无风险利率, σ 是股票价格波动参数, μ 是股票价格漂移系数, W 为标准 Brown 运动。这里假设不存在交易成本, 而且两种金融资产都能够自由地瞬时交易。

2. 零利率

在分析 Black-Scholes 模型时, 首先遇到的参数是利率 r, 它使得问题的处理变得比较繁琐。为简便计, 先考虑 $r = 0$。

现在要考虑的问题是: 对于任意的请求权 X, 它到时刻 T 时完全确定, 是否可以找到一个可复制策略 (ϕ, ψ)? 下面分三个步骤讨论。

(1) 寻找概率测度 $\mathbb{Q}$, 使得 S 是 $\mathbb{Q}$- 鞅。

(2) 构造过程 $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$。

(3) 寻找可预期过程 ϕ_t, 使得 $\mathrm{d}E_t = \phi_t \mathrm{d}S_t$。

第一步(应用Girsanov定理)

首先考虑 S 满足的随机微分方程, 然后应用 Girsanov 定理, 寻找测度 $\mathbb{Q}$, 使得 S 是 $\mathbb{Q}$- 鞅。

股票价格过程是一个指数 Brown 运动, $S_t = S_0 \exp(\sigma W_t + \mu t)$, 应用 Itô 公式, 可得随机微分方程

$$\mathrm{d}S_t = \sigma S_t \mathrm{d}W_t + (\mu + \frac{1}{2}\sigma^2) S_t \mathrm{d}t$$

为使 S 是鞅, 需要消去随机微分方程中的漂移项。设 $\gamma_t = \gamma = (\mu + \frac{1}{2}\sigma^2)/\sigma$ 为常数值过程, 应用 Girsanov 定理, 存在一个测度 $\mathbb{Q}$ 使得 $\widetilde{W}_t = W_t + \gamma t$ 是 $\mathbb{Q}$–Brown 运

动。代入上面的随机微分方程, 可得在测度 $\mathbb{Q}$ 下, S 满足以下随机微分方程

$$dS_t = \sigma S_t d\widetilde{W}_t$$

上式构造了一个指数鞅, 因此 S 是 $\mathbb{Q}$- 鞅。

第二步 给定 $\mathbb{Q}$, $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$ 就把 X 转化为一个 $\mathbb{Q}$- 鞅。

第三步(应用鞅表示定理)

E_t 和 S_t 都是 $\mathbb{Q}$- 鞅, 应用鞅表示定理, 存在一个可预期过程 ϕ, 使得 $E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$ 可由 S_t 表示出来。(为了应用鞅表示定理, 需要验证 S_t 的波动参数是否总取正值。实际上, 这里的波动参数是 σS_t, 而 σ 和 S_t 都取正值)。因此,

$$E_t = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t) = \mathbb{E}_{\mathbb{Q}}(X) + \int_0^t \phi_s dS_s$$

或者 $dE_t = \phi_t dS_t$。

3. 可复制策略

可选择的策略如下: 在时刻 t 时持有 ϕ_t 个单位的股票, 持有 $\psi_t = E_t - \phi_t S_t$ 债券。注意到 $B_t \equiv 1$, 在时刻 t 时投资组合的价值为

$$V_t = \phi_t S_t + \psi_t B_t = E_t$$

因此

$$dV_t = dE_t = \phi_t dS_t = \phi_t dS_t + \psi_t dB_t$$

从而 (ϕ, ψ) 是自融资的。另外, V 的终值是 $E_T = \mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_T) = X$, 因此 (ϕ, ψ) 是关于 X 的可复制策略。特别地, X 在时刻 0 时的价值为 $E_0 = \mathbb{E}_{\mathbb{Q}}(X)$。换言之, 在使股票价格过程 S 为鞅的测度 $\mathbb{Q}$ 下, 请求权 X 的价格是关于测度 $\mathbb{Q}$ 的期望。

上面的结论隐含了一些令人惊讶的事实。首先, 对于任意请求权都存在可复制策略。另外, 尽管请求权的终值依赖于股票价格过程, 它在开始时并不能确定, 但请求权仍能够在市场上交易。

其次, 衍生产品的价格有简单的表达式, 即可表示为请求权关于测度 $\mathbb{Q}$ 的期望。而在实际测度下, S 是具有漂移系数 μ 和波动参数 σ 的指数 Brown 运动。但实际测度无助于确定衍生证券的无套利价格。

最后, 在鞅测度 $\mathbb{Q}$ 下, 过程 S 的形式比较简单。对一个特定的请求权, 需要在鞅测度 $\mathbb{Q}$ 下计算价格。因为请求权依赖于 S, 因此, 在计算价格时, 通常要计算直到 $t = T$ 时关于 S 的某个函数的期望。如果在测度 $\mathbb{Q}$ 下, S 的形式比较复杂, 那么有关计算是很不方便的。实际上在测度 $\mathbb{Q}$ 下, S 也是一个指数 Brown 运动。求解随机微分方程, 可得

$$S_t = S_0 \exp\left(\sigma \widetilde{W}_t - \frac{1}{2}\sigma^2 t\right)$$

4. 非零利率

现在假设 $r \neq 0$。为消除利率的影响, 可应用贴现的方式。称 B_t^{-1} 为贴现过程, 构造贴现股票价格 $Z_t = B_t^{-1}S_t$, 贴现请求权 $B_T^{-1}X$。

对贴现股票价格过程 Z_t, 可以证明它满足如下随机微分方程

$$\mathrm{d}Z_t = Z_t\left(\sigma \mathrm{d}W_t + \left(\mu - r + \frac{1}{2}\sigma^2\right)\mathrm{d}t\right)$$

下面构造可复制策略的步骤和零利率情形下相应的步骤相同。

第一步 (**应用Girsanov定理**)

为把 Z_t 转化为鞅, 设 $\gamma_t = \gamma = \left(\mu - r + \frac{1}{2}\sigma^2\right)/\sigma$, 引入 $\widetilde{W}_t = W_t + \gamma t$。应用 Girsanov 定理, 存在一个与初始测度 $\mathbb{P}$ 等价的测度 $\mathbb{Q}$, 使得 $\widetilde{W}_t$ 为 $\mathbb{Q}$–Brown 运动, 而且

$$\mathrm{d}Z_t = \sigma Z_t \mathrm{d}\widetilde{W}_t$$

因此在测度 $\mathbb{Q}$ 下, Z 是鞅。

第二步 构造 $E_t = \mathbb{E}_{\mathbb{Q}}(B_T^{-1}X|\mathcal{F}_t)$, 它也是 $\mathbb{Q}$– 鞅。

第三步 (**应用鞅表示定理**)

贴现股票价格 Z 是一个 $\mathbb{Q}$– 鞅, 同时贴现请求权的条件期望过程 E 也是一个 $\mathbb{Q}$– 鞅。应用鞅表示定理, 存在可预期过程 ϕ, 使得 $\mathrm{d}E_t = \phi_t \mathrm{d}Z_t$。$\phi$ 给出了贴现请求权与贴现股票之间的关系, 可以设想请求权与股票之间的关系也应由 ϕ 给出。

在贴现情况下, 在时刻 t 时持有的债券数量是 $\psi_t = E_t - \phi_t Z_t$, 可以设想实际持有的债券的单位也是 ψ_t。

5. 可复制策略

可复制策略为: 在时刻 t 时持有 ϕ_t 个单位的股票, 持有 $\psi_t = E_t - \phi_t Z_t$ 个单位的债券。为说明这一点, 首先验证该策略是自融资的。

投资组合 (ϕ, ψ) 的价值为 $V_t = \phi_t S_t + \psi_t B_t = B_t E_t$。因此反复应用 Itô 公式, 即得

$$\begin{aligned}
\mathrm{d}V_t &= B_t \mathrm{d}E_t + E_t \mathrm{d}B_t \\
&= \phi_t B_t \mathrm{d}Z_t + (\phi_t Z_t + \psi_t)\mathrm{d}B_t \\
&= \phi_t \mathrm{d}(B_t Z_t) + \psi_t \mathrm{d}B_t \\
&= \phi_t \mathrm{d}S_t + \psi_t \mathrm{d}B_t
\end{aligned}$$

从而 (ϕ, ψ) 是自融资的。

其次, 投资策略在时刻 T 时复制了请求权 X。这可验证如下: 在时刻 T 时持有 ϕ_T 个单位的股票和 ψ_T 个单位的债券, 其价值是 $\phi_T S_T + \psi_T B_T = B_T E_T = X$。

有了以上准备，最后可得：请求权 X 在时刻 0 的初始价格为 $\phi_0 S_0 + \psi_0 B_0 = E_0 = \mathbb{E}_{\mathbb{Q}}(B_T^{-1}X)$。进一步，请求权 X 在时刻 t 时的无套利价格为

$$V_t = B_t \mathbb{E}_{\mathbb{Q}}(B_T^{-1}X|\mathcal{F}_t) = \mathrm{e}^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}(X|\mathcal{F}_t)$$

6. 看涨期权

现在给出一个例子，说明在 Black-Scholes 模型中，如何对看涨期权定价。设看涨期权的执行日期为 T，事先约定的执行价格为 k，那么请求权 $X = \max(S_T - k, 0)$，或记为 $(S_T - k)^+$。

首先求解 V_0，它是可复制资产在时刻 0 时的价值，前述结论表明

$$V_0 = \mathrm{e}^{-rT}\mathbb{E}_{\mathbb{Q}}((S_T - k)^+)$$

其中 $\mathbb{Q}$ 是 $\{B_t^{-1}S_t, t \geqslant 0\}$ 的鞅测度。

注意到请求权 $(S_T - k)^+$ 仅依赖于在执行日期 T 的股票价格，因此，为计算请求权的期望，只需考虑在测度 $\mathbb{Q}$ 下 S_T 的边际分布。

记时刻 0 时的股票价格 $S_0 = s$。在测度 $\mathbb{Q}$ 下，$S_t = s\exp\left(\sigma\widetilde{W}_t + \left(r - \frac{1}{2}\sigma^2\right)t\right)$。设 $Z \sim N\left(-\frac{1}{2}\sigma^2 T, \sigma^2 T\right)$，那么 $S_T = s\mathrm{e}^{(Z+rT)}$，从而

$$\begin{aligned} V_0 &= \mathrm{e}^{-rT}\mathbb{E}_{\mathbb{Q}}\left((s\mathrm{e}^{(Z+rT)} - k)^+\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2 T}}\int_{\log(k/s)-rT}^{\infty}(s\mathrm{e}^x - k\mathrm{e}^{-rT})\exp\left(-\frac{\left(x + \frac{1}{2}\sigma^2 T\right)^2}{2\sigma^2 T}\right)\mathrm{d}x \end{aligned}$$

上述积分可通过变量替换分解为两个标准正态积分。记 $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x}\exp(-y^2/2)\mathrm{d}y$，最后化简，得到 $V_0 = V(s,T)$ 的如下表示。

Black-Scholes公式

$$V(s,T) = s\Phi\left(\frac{\log(s/k) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) - k\mathrm{e}^{-rT}\Phi\left(\frac{\log(s/k) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right)$$

上式就是欧式看涨期权的 Black-Scholes 定价公式。

第十二章　平稳独立增量过程

§1　一些性质

1. 引言

平稳独立增量过程又称为 Lévy 过程，关于这类过程的系统研究起始于20世纪30 年代，法国概率论学者 P.Lévy 做出了开创性的贡献。至今，关于 Lévy 过程的理论及其应用的研究仍然是很活跃的专题。从理论上讲，Lévy 过程是一类重要的 Markov 过程，从而在理论研究中可以应用 Fourier 分析、位势理论、以及古典概率方法等工具。同时，Lévy 过程是一类很特殊的半鞅，从而在理论研究中可以应用随机微积分。从应用上讲，存储过程、保险风险过程等都是特殊的 Lévy 过程。当前，Lévy 过程在数理金融中也得到了重要应用，其中的原因之一就是很多无穷可分分布是厚尾的。

本章简要总结与 Lévy 过程有关的一些最常见性质与结论。

2. Markov 性

给定概率空间 $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geqslant 0},\mathbb{P})$，设 $(\mathcal{F}_t)_{t\geqslant 0}$ 是完备的、右连续的。设取值于 $\mathbb{R}^d$ 的随机过程 $X=\{X_t,t\geqslant 0\}$ 是可适应的。如果对任意 $t,s\geqslant 0$，$X_{t+s}-X_t$ 与 $\mathcal{F}_t$ 独立，而且 $X_{t+s}-X_t$ 与 X_s 同分布 (从而 $X_0=0$)，另外假设 X 的样本轨道是右连续的，而且左极限存在，那么称过程 X 为 Lévy 过程。

由以上定义可知，对 $t\geqslant 0$，如记 $X'_s=X_{t+s}$，那么在条件 $X_t=x$ 下，过程 $X'=\{X'_s,s\geqslant 0\}$ 与 $\mathcal{F}_t$ 独立，而且 $X'=\{X'_s,s\geqslant 0\}$ 与 $x+X=\{x+X_s,s\geqslant 0\}$ 同

分布。这是很特殊的 Markov 性。

为方便记, 对 $x \in \mathbb{R}^d$, 以 $\mathbb{P}^x$ 表示平移过程 $x+X$ 的分布。

下面引入与 Markov 性相关联的两个线性算子。

(i) 半群算子

对有界非负可测函数 $f:\mathbb{R}^d \to \mathbb{R}$, 半群算子 $\{P_t, t \geqslant 0\}$ 定义如下

$$P_t f(x) = \mathbb{E}_x[f(X_t)] = \mathbb{E}[f(x+X_t)],\ x \in \mathbb{R}^d$$

在上述定义中, $P_t f$ 是一个函数。由 Markov 性与关于条件期望的塔式性质, 可得如下半群性质

$$P_t \circ P_s = P_{t+s}$$

其中符号"$\circ$"表示复合运算。

(ii) 预解算子

预解算子 $\{U^q, q>0\}$ 定义如下

$$U^q f(x) = \int_0^\infty \mathrm{e}^{-qt} P_t f(x)\mathrm{d}t = \mathbb{E}\left(\int_0^\infty \mathrm{e}^{-qt} f(x+X_t)\mathrm{d}t\right)$$

由 $\{P_t, t \geqslant 0\}$ 的半群性质, 可得如下预解方程: 对 $q, r>0$

$$U^q f - U^r f + (q-r)U^q(U^r f) = 0$$

3. 无穷可分分布与 Lévy-Khintchine 公式

对给定的 Lévy 过程 X, 首先对任意正整数 n,

$$X_1 = X_{1/n} + (X_{2/n} - X_{1/n}) + \cdots + (X_{n/n} - X_{(n-1)/n})$$

由平稳独立增量性可知, X_1 可分解为 n 个独立同分布的随机变量之和, 从而 Lévy 过程的一维分布是无穷可分的。反之, 可以证明每个无穷可分分布都可以视为某个 Lévy 过程在时刻 1 时的变量分布。

引用无穷可分分布的结论, X_1 的特征函数有如下形式

$$\mathbb{E}\left(\mathrm{e}^{i\langle\lambda, X_1\rangle}\right) = \mathrm{e}^{-\Psi(\lambda)}, \lambda \in \mathbb{R}^d$$

其中 $\langle\cdot,\cdot\rangle$ 表示 $\mathbb{R}^d$ 空间中的内积, $\Psi:\mathbb{R}^d \to \mathbb{C}$ 是连续函数, 而且 $\Psi(0)=0$, 称 Ψ 为特征指数。无穷可分分布的 Ψ 有如下形式, 称为 Lévy-Khintchine 公式。

$$\Psi(\lambda) = i\langle a, \lambda\rangle + \frac{1}{2}Q(\lambda) + \int_{\mathbb{R}^d}(1 - \mathrm{e}^{i\langle\lambda, x\rangle} + i\langle\lambda, x\rangle \mathbf{1}_{\{|x|<1\}})\Pi(\mathrm{d}x)$$

其中 $a \in \mathbb{R}^d$, Q 是 $\mathbb{R}^d$ 上半正定二次型, Π 是 $\mathbb{R}^d\backslash\{0\}$ 上的测度, 称为 Lévy 测度, 它满足 $\int(1 \wedge |x|^2)\Pi(\mathrm{d}x) < \infty$。另外, a, Q 和 Π 由 Ψ 唯一确定。

下面给出 $\mathbb{R}$ 上一些常见的无穷可分分布, 相应地就有 Lévy 过程。

例 1 Poisson 分布。设参数为 c, 那么该分布的特征函数为

$$\sum_{n=0}^{\infty} \frac{c^n}{n!} \mathrm{e}^{-c+i\lambda n} = \exp\{-c(1-\mathrm{e}^{i\lambda})\}$$

此时特征指数为

$$\Psi(\lambda) = c(1-\mathrm{e}^{i\lambda})$$

与一般的 Lévy-Khintchine 公式做比较, 可知 Lévy 测度为 $c\delta_1$, 该测度只在 1 点有正测度 c。相应的 Lévy 过程就是参数为 c 的 Poisson 过程。

例 2 标准正态分布。特征函数为

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda x} \mathrm{e}^{-x^2/2} \mathrm{d}x = \exp\{-\lambda^2/2\}$$

特征指数为

$$\Psi(\lambda) = \lambda^2/2$$

与一般的 Lévy-Khintchine 公式做比较, 可知 Lévy 测度为 0。相应的 Lévy 过程就是标准 Brown 运动。

例 3 Cauchy 分布。特征函数为

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda x} \frac{1}{1+x^2} \mathrm{d}x = \exp\{-|\lambda|\} = \exp\left\{-\frac{1}{\pi} \int_{-\infty}^{\infty} (1-\mathrm{e}^{i\lambda x}) x^{-2} \mathrm{d}x\right\}$$

特征指数为

$$\Psi(\lambda) = |\lambda|$$

与一般的 Lévy-Khintchine 公式做比较, 可知 Lévy 测度为 $\pi^{-1}x^{-2}\mathrm{d}x$。相应的 Lévy 过程就是标准 Cauchy 过程。

例 4 Gamma 分布。设分布参数为 c, 该分布的特征函数为

$$\frac{1}{\Gamma(c)} \int_0^{\infty} \mathrm{e}^{i\lambda x} x^{c-1} \mathrm{e}^{-x} \mathrm{d}x = (1-i\lambda)^{-c} = \exp\left\{-c \int_0^{\infty} (1-\mathrm{e}^{i\lambda x}) \mathrm{e}^{-x} x^{-1} \mathrm{d}x\right\}$$

特征指数为

$$\Psi(\lambda) = c \int_0^{\infty} (1-\mathrm{e}^{i\lambda x}) \mathrm{e}^{-x} x^{-1} \mathrm{d}x$$

与一般的 Lévy-Khintchine 公式做比较, 可知 Lévy 测度为 $c\mathbf{1}_{\{x>0\}} x^{-1} \mathrm{e}^{-x} \mathrm{d}x$。相应的 Lévy 过程就是 Gamma 过程。

例 5 严格稳定分布。这类分布是正态分布的推广, 特别地它们出现于关于独立同分布、而方差为无穷大的随机变量之和的极限定理中。这类分布依赖于三个参数, 其中最重要的是指标参数 $\alpha \in (0,2]$。当 $\alpha = 2$ 时, 稳定分布就是正态分布。当 $\alpha = 1$ 时, 稳定分布就是 Cauchy 分布的变换。因此, 通常只限于考虑 $\alpha \in (0,1) \cup (1,2)$。另外两个参数是斜度参数 $\beta \in [-1,1]$ 与标度参数 $\gamma > 0$。参数为 (α,β,γ) 的严格稳定分布的特征函数有如下形式

$$\exp\{\gamma|\lambda|^{\alpha}[1 - i\beta \mathrm{sgn}(\lambda)\tan(\pi\alpha/2)]\}$$

在上式中, $\mathrm{sgn}(\lambda)$ 表示符号函数, 即

$$\mathrm{sgn}(\lambda) = \begin{cases} 1, & \lambda > 0 \\ 0, & \lambda = 0 \\ -1, & \lambda < 0 \end{cases}$$

可以证明 Lévy 测度为

$$\Pi(\mathrm{d}x) = \begin{cases} c^{+}x^{-\alpha-1}\mathrm{d}x, & x > 0 \\ c^{-}|x|^{-\alpha-1}\mathrm{d}x, & x < 0 \end{cases}$$

其中 c^{+} 和 c^{-} 是两个非负实数, 使得 $\beta = (c^{+} - c^{-})/(c^{+} + c^{-})$。相应的 Lévy 过程称为指标为 α, 斜度为 β 的稳定 Lévy 过程。

关于稳定分布, 下面给出一个定义。

定义 对随机变量 X, 如果对任意 $n \geqslant 2$, 存在实数 $c_n > 0$ 和 d_n, 使得 $X_1 + X_2 + \cdots + X_n$ 与 $c_n X + d_n$ 同分布, 其中 $X_1, X_2, \cdots$ 相互独立, 并与 X 同分布, 那么称 X 的分布为稳定分布。如引入符号 $S_n = X_1 + X_2 + \cdots + X_n$, 那么上述条件表明

$$\frac{S_n - d_n}{c_n} = X$$

定义表明 S_n 的分布本质上也是稳定分布。实际上, 可以证明, $c_n = n^{1/\alpha}$, $\alpha \in (0,2]$。当 $\alpha \in (0,2)$ 时, 稳定分布的概率密度函数收敛到 0 的速度较慢。当 $x \to \infty$ 时, 概率密度函数趋近于 $kx^{-\alpha}$, k 是某常数。

对严格稳定过程, 容易验证自相似性成立, 即如果 X 是指标为 α 的严格稳定过程, 那么对 $k > 0$, $\{k^{-1/\alpha}X_{kt}\}$ 与 $\{X_t\}$ 同分布。这等价于特征指数所满足的关系 $\psi(k\lambda) = k^{\alpha}\psi(\lambda)$。

再次应用平稳独立增量性, 可知当 t 是有理数时, X_t 的特征指数为 $t\Psi$。进一步, 由样本轨道的右连续性, 可得对任意正数 t, X_t 的特征指数也为 $t\Psi$, 即对任意

正数 t,

$$\mathbb{E}\left(e^{i\langle\lambda,X_t\rangle}\right)=e^{-t\Psi(\lambda)}$$

因此特征指数确定了 X_t 的分布, 由平稳独立增量性, Lévy 过程的分布也就完全由特征指数确定。特别地, 半群算子与预解算子也就完全由特征指数通过 Fourier 变换来确定。

具体地讲, 对可积函数 g, 记它的 Fourier 变换为 $\mathcal{F}(g)$,

$$\mathcal{F}(g)(\xi)=\int_{\mathbb{R}^d}e^{i\langle x,\xi\rangle}g(x)\mathrm{d}x,\xi\in\mathbb{R}^d$$

那么 P_tf 的 Fourier 变换 $\mathcal{F}(P_tf)$ 可表示如下

$$\mathcal{F}(P_tf)(\xi)=e^{-t\Psi(-\xi)}\mathcal{F}(f)(\xi),\xi\in\mathbb{R}^d \tag{1}$$

类似地, U^qf 的 Fourier 变换 $\mathcal{F}(U^qf)$ 可表示如下

$$\mathcal{F}(U^qf)(\xi)=\frac{\mathcal{F}(f)(\xi)}{q+\Psi(-\xi)},\xi\in\mathbb{R}^d \tag{2}$$

下面只对 (1) 式给出证明。

$$\begin{aligned}\mathcal{F}(P_tf)(\xi)&=\int_{\mathbb{R}^d}e^{i\langle x,\xi\rangle}P_tf(x)\mathrm{d}x\\&=\int_{\mathbb{R}^d}e^{i\langle x,\xi\rangle}\mathbb{E}[f(x+X_t)]\mathrm{d}x\\&=\mathbb{E}\left(\int_{\mathbb{R}^d}e^{i\langle x,\xi\rangle}f(x+X_t)\mathrm{d}x\right)\\&=\mathbb{E}\left(\int_{\mathbb{R}^d}e^{i\langle y-X_t,\xi\rangle}f(y)\mathrm{d}y\right)\\&=\mathbb{E}\left(e^{i\langle-\xi,X_t\rangle}\right)\int_{\mathbb{R}^d}e^{i\langle y,\xi\rangle}f(y)\mathrm{d}y\\&=e^{-t\Psi(-\xi)}\mathcal{F}(f)(\xi)\end{aligned}$$

4. 一维 Lévy 过程

当 $d=1$ 时, 在一些特殊情形下, 应用 Laplace 变换研究 Lévy 过程的分布更方便。对 $\lambda>0$, 记

$$\Phi(\lambda)=\psi(i\lambda)=-a\lambda-\frac{Q}{2}\lambda^2+\int_{\mathbb{R}}\left(1-e^{-\lambda x}-\lambda x\mathbf{1}_{\{|x|<1\}}\right)\Pi(\mathrm{d}x)$$

称 $\Phi:[0,\infty)\to\mathbb{R}$ 为 Laplace 指数。对 $t\geqslant0$

$$\mathbb{E}\left(e^{-\lambda X_t}\right)=\exp\{-t\Phi(\lambda)\}$$

下面简要介绍两类特殊的 Lévy 过程。

(i) 谱正的 Lévy 过程 X, 此时在 X 的每个不连续点的跳跃度都是正的。传统的保险风险过程就是这类过程。

(ii) 从属过程, 此时 Lévy 过程 X 的样本轨道是单增的。复合 Poisson 过程就是很特殊的从属过程。对这类过程, 可以验证 $a \leqslant 0$, $Q = 0$, Π 是 $(0,\infty)$ 上的测度, 满足条件 $\int_{(0,\infty)} (1 \wedge x)\Pi(\mathrm{d}x) < \infty$。另外, Laplace 指数可记为如下形式

$$\Phi(\lambda) = c\lambda + \int_{\mathbb{R}} (1 - \mathrm{e}^{-\lambda x})\Pi(\mathrm{d}x)$$

其中 $c \geqslant 0$。

对指标为 $\alpha \in (0,1)$ 的严格稳定从属过程, Lévy 测度为 $\Lambda(\mathrm{d}x) = x^{-1-\alpha}\mathrm{d}x$, 调整单位后, Laplace 指数为

$$\frac{\alpha}{\Gamma(1-\alpha)} \int_0^\infty (1 - \mathrm{e}^{-\lambda x}) x^{-1-\alpha} \mathrm{d}x = \lambda^\alpha$$

对 Gamma 过程, Laplace 指数为

$$\psi(\lambda) = a\log\left(1 + \frac{\lambda}{b}\right)$$

应用 Frullani 积分

$$\log\left(1 + \frac{\lambda}{b}\right) = \int_0^\infty (1 - \mathrm{e}^{-\lambda x}) x^{-1} \mathrm{e}^{-bx} \mathrm{d}x$$

可知 Gamma 过程对应的 Lévy 测度为 $\Lambda(\mathrm{d}x) = ax^{-1}\mathrm{e}^{-bx}\mathrm{d}x$。

§2 Lévy 过程的结构

本节给出 Lévy-Khintchine 公式的概率解释。为研究 Lévy 过程 X 的结构, 要从 X 的跳跃过程出发。

1. Poisson 点过程

对 $t > 0$, 记 $\Delta X_t = X_t - X_{t-}$, 对 Borel 集合 $B \subseteq \mathbb{R}^d$, 当 B 与原点的距离大于 0 时, 记 $N_t^B = \#\{s \in (0,t] : \Delta X_s \in B\}$ 为 X 到时刻 t 时跳跃度属于 B 的跳跃点的个数, N_t^B 是有限的。由 X 的平稳独立增量性, 可知 $N^B = \{N_t^B, t \geqslant 0\}$ 也有平稳独立增量性。注意到 N^B 是计数过程, 而且每次的跳跃度只能是 1。在随机过程理论中已证明具备这些性质的 N^B 是 Poisson 过程。

另外，如果 $B_1, B_2, \cdots$ 是不相交的 Borel 集，那么计数过程 $N^{B_1}, N^{B_2}, \cdots$ 都是 Poisson 过程，而且任何两个没有公共跳跃点。由 Poisson 过程的结论，可得 $N^{B_1}, N^{B_2}, \cdots$ 是独立的。

设 Poisson 过程 N^B 的强度参数为 $\Lambda(B)$，考虑 B 的可列分割 $B_1, B_2, \cdots$，那么

$$N^B = N^{B_1} + N^{B_2} + \cdots$$

上式右边是独立的 Poisson 过程之和，所以它也是 Poisson 过程，强度参数为 $\Lambda(B_1)+\Lambda(B_2)+\cdots$，由此即得 $\Lambda(B)=\Lambda(B_1)+\Lambda(B_2)+\cdots$。总之 Λ 是 $\mathbb{R}^d-\{0\}$ 上的测度，在原点临域之外的集合的测度是有限的。引用随机过程的术语就是，Lévy 过程的跳跃过程是特征测度为 Λ 的 Poisson 点过程。这里的 Λ 就是 Lévy 过程 X 的 Lévy 测度 Π。

作为一个例子，考虑复合 Poisson 过程。设 Λ 是 $\mathbb{R}^d-\{0\}$ 上的有限测度，$\Delta=\{\Delta_t, t\geqslant 0\}$ 是特征测度为 Λ 的 Poisson 点过程。因为 Λ 是有限的，Δ 在任意有界时间区间上只有有限次跳跃，从而 $Y_t=\sum\limits_{0\leqslant s\leqslant t}\Delta_s$ 是有定义的。Y 是右连续的，而且 $\Delta Y=\Delta$。Y 是 Lévy 过程，为确定它的分布，可计算特征函数。由 Poisson 点过程的指数公式，可得

$$\mathbb{E}\left(\mathrm{e}^{i\langle\lambda,Y_t\rangle}\right)=\mathbb{E}\left(\exp\left\{i\sum_{0\leqslant s\leqslant t}\langle\lambda,\Delta_s\rangle\right\}\right)=\exp\left\{-t\int_{\mathbb{R}^d}(1-\mathrm{e}^{i\langle\lambda,x\rangle})\Lambda(\mathrm{d}x)\right\}$$

复合 Poisson 过程又可记为

$$X_t=\sum_{i=1}^{N_t}\xi_i$$

其中 ξ_i 为独立同分布随机变量，记分布函数为 v, $v(0)=0$, N 为参数为 λ 的 Poisson 过程，N 与 ξ_i 独立。此时 X 的特征指数为 $\Psi(\lambda)=a\int_{\mathbb{R}^d}(1-\mathrm{e}^{i\langle\lambda,x\rangle})v(\mathrm{d}x)$，$a>0$ 为某常数。

2. Lévy 过程的分解

对一般的 Lévy 过程，特征指数 Ψ 有以下分解

$$\Psi=\Psi^{(0)}+\Psi^{(1)}+\Psi^{(2)}+\Psi^{(3)}.$$

其中

$$\Psi^{(0)}(\lambda) = i\langle a, \lambda\rangle, \quad \Psi^{(1)}(\lambda) = \frac{1}{2}Q(\lambda)$$
$$\Psi^{(2)}(\lambda) = \int_{\mathbb{R}^d} (1 - \mathrm{e}^{i\langle\lambda,x\rangle})\mathbf{1}_{\{|x|\geqslant 1\}}\Pi(\mathrm{d}x)$$
$$\Psi^{(3)}(\lambda) = \int_{\mathbb{R}^d} \left(1 - \mathrm{e}^{i\langle\lambda,x\rangle} + i\langle\lambda, x\rangle\right)\mathbf{1}_{\{|x|<1\}}\Pi(\mathrm{d}x)$$

$\Psi^{(i)}$ 都是特征指数, $\Psi^{(0)}$ 对应于确定性的线性过程 ct, $\Psi^{(1)}$ 对应于 d 维 Brown 运动的线性变换, 上面的例子表明 $\Psi^{(2)}$ 是 Lévy 测度为 $\mathbf{1}_{\{|x|\geqslant 1\}}\Pi(\mathrm{d}x)$ 的复合 Poisson 过程的特征指数。$\Psi^{(3)}$ 对应的 Lévy 过程比较复杂, 它是跳跃度小于 1 的纯跳跃过程。

由以上讨论, 一般的 Lévy 过程 X 可以表示为四个独立的 Lévy 过程之和, $X = Y^{(0)} + Y^{(1)} + Y^{(2)} + Y^{(3)}$, 其中 $Y^{(0)}$ 为确定性的线性过程 ct, $Y^{(1)}$ 为 d 维 Brown 运动的线性变换, $Y^{(2)}$ 为跳跃度大于等于 1 的复合 Poisson 过程, $Y^{(3)}$ 为跳跃度小于 1 的纯跳跃过程。

上述结论称为 Lévy 过程的 Lévy-Itô分解。

§3 Feynman-Kac 公式

1. Feynman-Kac 公式

设 $f : \mathbb{R}^d \to [0, \infty)$ 是可测函数, 定义

$$A_f(t) = \int_0^t f(X_s)\mathrm{d}s \tag{3}$$

$A_f(t)$ 是单增的可加泛函。Feynman-Kac 公式是研究关于 $A_f(t)$ 的信息的工具之一, 它与 Laplace 变换有关。

具体地讲, 设 $g : \mathbb{R}^d \to [0, \infty)$ 是另一可测函数, 对 $q > 0$, $x \in \mathbb{R}^d$, 定义

$$V_f^q g(x) = \mathbb{E}_x\left(\int_0^\infty \mathrm{e}^{-qt} g(X_t)\mathrm{e}^{-A_f(t)}\mathrm{d}t\right) \tag{4}$$

定理 1 (Feynman-Kac 公式) 对一般的 Lévy 过程, 下面的关系成立

$$U^q\left(fV_f^q g\right) = U^q g - V_f^q g$$

证明　直接由定义可得

$$
\begin{aligned}
U^q\left(fV_f^q g\right)(x) &= \mathbb{E}_x\left(\int_0^\infty \mathrm{d}t\mathrm{e}^{-qt}f(X_t)\mathbb{E}_{X_t}\left(\int_0^\infty \mathrm{d}s\mathrm{e}^{-qs}g(X_s)\mathrm{e}^{-A_f(s)}\right)\right)\\
&= \mathbb{E}_x\left(\int_0^\infty \mathrm{d}t\mathrm{e}^{-qt}f(X_t)\mathbb{E}_x\left(\int_0^\infty \mathrm{d}s\mathrm{e}^{-qs}g(X_{s+t})\mathrm{e}^{-(A_f(s+t)-A_f(t))}\Big|\mathcal{F}_t\right)\right)\\
&= \mathbb{E}_x\left(\int_0^\infty \mathrm{d}t\mathrm{e}^{-qt}f(X_t)\int_0^\infty \mathrm{d}s\mathrm{e}^{-qs}g(X_{s+t})\mathrm{e}^{A_f(t)-A_f(s+t)}\right)\\
&= \mathbb{E}_x\left(\int_0^\infty \mathrm{d}t\mathrm{e}^{-qt}f(X_t)\int_t^\infty \mathrm{d}s\mathrm{e}^{-q(s-t)}g(X_s)\mathrm{e}^{A_f(t)-A_f(s)}\right)\\
&= \mathbb{E}_x\left(\int_0^\infty \mathrm{d}s\mathrm{e}^{-qs}g(X_s)\int_0^s \mathrm{d}t\mathrm{e}^{A_f(t)-A_f(s)}f(X_t)\right)\\
&= \mathbb{E}_x\left(\int_0^\infty \mathrm{d}s\mathrm{e}^{-qs}g(X_s)\left(1-\mathrm{e}^{-A_f(s)}\right)\right)\\
&= U^q g(x) - V_f^q g(x)
\end{aligned}
$$

2．Feynman-Kac 公式与偏微分方程的联系

(i) 历史上, Feynman-Kac 公式的含义是指对一些微分方程, 它们的解可用随机过程给出概率表示, 这种含义和上面介绍的含义有所不同。

考虑如下抛物型方程

$$
\frac{\partial u}{\partial t} + ku = \frac{1}{2}\Delta u + g,\quad (t,x)\in(0,\infty)\times\mathbb{R}^d
$$

设初值条件为

$$
u(0,x) = f(x),\quad x\in\mathbb{R}^d
$$

其中函数 $k:\mathbb{R}^d\to[0,\infty), g:(0,\infty)\times\mathbb{R}^d\to\mathbb{R},\ f:\mathbb{R}^d\to\mathbb{R}$。

当 $g=0$ 时, 定义

$$
z_\alpha(x) = \int_0^\infty \mathrm{e}^{-\alpha t}u(t,x)\mathrm{d}t;\quad x\in\mathbb{R}^d
$$

当 $\alpha>0, x\in\mathbb{R}^d$ 时, 假设 $\lim\limits_{t\to\infty}\mathrm{e}^{-\alpha t}u(t,x)=0$, 那么应用分部积分, 形式上可得 Z_α 应满足如下椭圆型方程

$$
\frac{1}{2}\Delta z_\alpha = \frac{1}{2}\int_0^\infty \mathrm{e}^{-\alpha t}\Delta u\mathrm{d}t = (\alpha+k)z_\alpha - f
$$

上述方程变形就是

$$
(\alpha+k)z_\alpha = \frac{1}{2}\Delta z_\alpha + f \tag{5}
$$

方程 (5) 的解的随机表示记为 z, 它有如下形式

$$
\begin{aligned}
z(x) &= \mathbb{E}_x \int_0^\infty f(W_t)\exp\left\{-\alpha t - \int_0^t k(W_s)\mathrm{d}s\right\}\mathrm{d}t \\
&= \mathbb{E}_x\left(\int_0^\infty \mathrm{e}^{-\alpha t} f(W_t)\mathrm{e}^{-A_k(t)}\mathrm{d}t\right) \\
&= V_k^\alpha f(x)
\end{aligned} \tag{6}
$$

其中 W 是标准 Brown 运动。与 $V_k^\alpha f(x)$ 类似的符号已经在 (4) 式中出现。

一般地, 当 f 和 k 都是连续函数时, 设方程 (5) 的解为 ψ, 那么可以证明 $\psi \geqslant z$。

(ii) 在一维情形下, 方程 (5) 式变为

$$
(\alpha + k)z = \frac{1}{2}z'' + f \tag{7}
$$

此时, 如果进一步假设函数 f 满足以下条件

$$
\mathbb{E}_x \int_0^\infty \mathrm{e}^{-\alpha t}\left|f(W_t)\right|\mathrm{d}t < \infty,\ x \in \mathbb{R} \tag{8}
$$

那么可以证明随机表示 (6) 是方程 (7) 的解。

为证明上述结论, 注意到对一维的 Brown 运动 W, 预解算子有以下具体形式

$$
\begin{aligned}
U^\alpha g(x) &= \mathbb{E}_x \int_0^\infty \mathrm{e}^{-\alpha t} g(W_t)\mathrm{d}t \\
&= \frac{1}{\sqrt{2\alpha}} \int_{-\infty}^\infty \mathrm{e}^{-|y-x|\sqrt{2\alpha}} g(y)\mathrm{d}y \\
&= \frac{1}{\sqrt{2\alpha}}\left[\int_{-\infty}^x \mathrm{e}^{(y-x)\sqrt{2\alpha}} g(y)\mathrm{d}y + \int_x^\infty \mathrm{e}^{(x-y)\sqrt{2\alpha}} g(y)\mathrm{d}y\right]
\end{aligned}
$$

微分后, 得到

$$
\begin{aligned}
(U^\alpha g)'(x) &= \int_x^\infty \mathrm{e}^{(x-y)\sqrt{2\alpha}} g(y)\mathrm{d}y - \int_{-\infty}^x \mathrm{e}^{(y-x)\sqrt{2\alpha}} g(y)\mathrm{d}y \\
(U^\alpha g)''(x) &= -2g(x) + 2\alpha U^\alpha g(x)
\end{aligned}
$$

在上面的最后表达式中, 先后取 $g = kz$ 和 $g = f$, 得到两个表达式后, 两式相减得到一个表达式。注意到应用定理 1, 可得下式成立

$$
U^\alpha(kz) = U^\alpha(f) - z
$$

最后就得到随机表示 (6) 是方程 (7) 的解这一结论。

顺便指出, 由于

$$
\int_0^\infty \mathrm{e}^{-\alpha t}\frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-\xi^2/2t}\mathrm{d}t = \frac{1}{\sqrt{2\alpha}}\mathrm{e}^{-|\xi|\sqrt{2\alpha}},\ \alpha > 0,\ \xi \in \mathbb{R}
$$

(8) 式等价于以下条件

$$\int_{-\infty}^{\infty}|f(x+y)|\mathrm{e}^{-|y|\sqrt{2\alpha}}\mathrm{d}y<\infty,\ x\in\mathbb{R}$$

3. 微分方程的概率表示的应用例子：Arcsine 律

对标准 Brown 运动 W, 考虑在时刻 t 前 Brown 运动样本轨道处于上半平面的时间, 称它为占位时间, 记为 $\Gamma_+(t)=\int_0^t 1_{(0,\infty)}(W_s)\mathrm{d}s$。下面的定理表明 $\Gamma_+(t)$ 的分布是 Arcsine 律, 它是概率论中著名的分布之一。

定理 2 $\Gamma_+(t)$ 的分布函数为

$$\mathbb{P}[\Gamma_+(t)\leqslant\theta]=\int_0^{\theta/t}\frac{\mathrm{d}s}{\pi\sqrt{s(1-s)}}=\frac{2}{\pi}\arcsin\sqrt{\frac{\theta}{t}},\ 0\leqslant\theta\leqslant t \tag{9}$$

证明 对 $\alpha>0,\beta>0$, 引入函数

$$z(x)=\mathbb{E}_x\int_0^{\infty}\exp\left(-\alpha t-\beta\int_0^t 1_{(0,\infty)}(W_s)\mathrm{d}s\right)\mathrm{d}t \tag{10}$$

比较 (6) 式和 (10) 式, 即知在 (10) 式中 $k=\beta 1_{(0,\infty)},f=1$。相应地, 微分方程 (7) 式变为

$$\begin{aligned}\alpha z(x)&=\frac{1}{2}z''(x)-\beta z(x)+1,\quad x>0\\ \alpha z(x)&=\frac{1}{2}z''(x)+1,\quad x<0\end{aligned} \tag{11}$$

另外, 连续性条件为

$$z(0_+)=z(0_-);z'(0_+)=z'(0_-) \tag{12}$$

方程 (11) 的唯一有界解为

$$z(x)=\begin{cases}A\mathrm{e}^{-x\sqrt{2(\alpha+\beta)}}+\dfrac{1}{\alpha+\beta}, & x>0\\ B\mathrm{e}^{x\sqrt{2\alpha}}+\dfrac{1}{\alpha}, & x<0\end{cases}$$

再由 (12) 式确定系数 A 和 B, 最后得到

$$z(0)=\mathbb{E}\int_0^{\infty}\mathrm{e}^{-\alpha t}\mathrm{e}^{-\beta\Gamma_+(t)}\mathrm{d}t=\int_0^{\infty}\mathrm{e}^{-\alpha t}\mathbb{E}[\mathrm{e}^{-\beta\Gamma_+(t)}]\mathrm{d}t=\frac{1}{\sqrt{\alpha(\alpha+\beta)}} \tag{13}$$

为得到 (9) 式, 需要对 (13) 式反演。以下仅验证 (9)。

$$\begin{aligned}\int_0^\infty e^{-\alpha t}\int_0^t \frac{e^{-\beta\theta}}{\pi\sqrt{\theta(t-\theta)}}d\theta dt &= \int_0^\infty \frac{e^{-\beta\theta}}{\pi\sqrt{\theta}}\int_\theta^\infty \frac{e^{-\alpha t}}{\sqrt{t-\theta}}dt d\theta \\ &= \frac{1}{\pi}\int_0^\infty \frac{e^{-(\alpha+\beta)\theta}}{\sqrt{\theta}}\int_0^\infty \frac{e^{-\alpha t}}{\sqrt{s}}dt d\theta \\ &= \frac{1}{\sqrt{\alpha(\alpha+\beta)}}\end{aligned} \tag{14}$$

在上式推导中用到了如下积分, 它可通过 Gamma 函数 $\Gamma(s)$ 变形得到,

$$\int_0^\infty \frac{e^{-\gamma t}}{\sqrt{t}}dt = \sqrt{\frac{\pi}{\gamma}},\ \gamma > 0$$

比较 (13) 式和 (14) 式, 即得

$$\mathbb{E}e^{-\beta\Gamma_+(t)} = \int_0^t \frac{e^{-\beta\theta}}{\pi\sqrt{\theta(t-\theta)}}d\theta \tag{15}$$

设 $\Gamma_+(t)$ 的密度函数为 $h(\theta)$, 那么由 (15) 式,

$$h(\theta) = \frac{1}{\pi\sqrt{\theta(t-\theta)}}$$

最后 $\Gamma_+(t)$ 的分布函数为

$$\mathbb{P}[\Gamma_+(t) \leqslant \theta] = \int_0^\theta \frac{1}{\pi\sqrt{r(t-r)}}dr = \int_0^{\theta/t} \frac{ds}{\pi\sqrt{s(1-s)}} = \frac{2}{\pi}\arcsin\sqrt{\frac{\theta}{t}}$$

从而定理结论成立。



第十三章 更新过程

§1 基本概念

1. 定义

更新过程是 Poisson 过程的推广。在 Poisson 过程中, 两个相邻跳跃点之间的时间 $\{t_i, i \geqslant 1\}$ 是独立同分布的指数随机变量。在一般的更新理论中, t_i 的分布不限于指数分布, 而且 $\{t_i, i \geqslant 1\}$ 的分布可以不同。在通常的更新过程中, 假设 $\{t_i, i \geqslant 1\}$ 是独立同分布的。

设想在时刻 0 时安装了一只新灯管, 经过 t_1 段时间灯管失效, 此时更换灯管, 又经过 t_2 段时间灯管失效, 再更换灯管, 依次类推, 这就给出了一个具体的更新过程。记 $T_n = t_1 + \cdots + t_n$, 它就是前 n 只灯管的总寿命。而到时刻 t 时, 已经更换的总的灯管数可表示为

$$N(t) = \max\{n : T_n \leqslant t\}$$

更新理论在很多实际问题中都有应用。下面给出两个例子。

例 1 (机器维修) 假设机器工作 s_i 段时间后就出现故障, 此时需要 u_i 段时间维修。记 $t_i = s_i + u_i$ 为一个工作与维修周期的时间。在实际问题中, 可以近似假设 t_i 是独立同分布的。

例 2 (Markov 链) 设 X 是离散时间的 Markov 链, $X_0 = i$。记 T_n 为 X 第 n 次回到初始状态 i 的时间, 那么应用 Markov 性, $\{t_n = T_n - T_{n-1}, n = 1, 2, \cdots\}$ 是独立同分布的, 从而 $\{t_n, n = 1, 2, \cdots\}$ 构成一个更新过程。

2. 计数过程 $N(t)$ 的期望

更新过程可以用更新时刻 T_n 来描述, 也可以用计数过程 $N(t)$ 来描述, 它们之间的关系为

$$\{N(t) \geqslant n\} = \{T_n \leqslant t\} \tag{1}$$

为计算 $\mathbb{E}[N(t)]$, 引入以下引理。

引理 1 如果 X 是非负随机变量, 那么 $\mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\}\mathrm{d}t$。

证明 引入示性函数

$$1_X(t) = \begin{cases} 1, & 0 < t < X \\ 0, & \text{其他} \end{cases}$$

显然, $\int_0^\infty 1_X(t)\mathrm{d}t = \int_0^X 1\mathrm{d}t = X, \mathbb{E}[1_X(t)] = \mathbb{P}\{X > t\}$, 从而

$$\mathbb{E}[X] = \mathbb{E}\int_0^X 1\mathrm{d}t = \int_0^\infty \mathbb{E}[1_X(t)]\mathrm{d}t = \int_0^\infty \mathbb{P}\{X > t\}\mathrm{d}t$$

引理 2 如果 X 是非负整数值随机变量, 那么

$$\mathbb{E}[X] = \sum_{n=1}^\infty \mathbb{P}\{X \geqslant n\} \tag{2}$$

证明 因为 X 取非负整数值, 所以当 $n-1 \leqslant t < n$ 时, $\{X > t\} = \{X \geqslant n\}$。因此

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\}\mathrm{d}t = \sum_{n=1}^\infty \int_{n-1}^n \mathbb{P}\{X > t\}\mathrm{d}t = \sum_{n=1}^\infty \mathbb{P}\{X \geqslant n\}$$

应用引理 2 的方法, 可以证明以下引理。

引理 3 设 H 为增函数, $H(0) = 0$, $h(x) = H'(x)$。如果 $X \geqslant 0$, 那么

$$\mathbb{E}[H(X)] = \int_0^\infty h(t)\mathbb{P}\{X > t\}\mathrm{d}t$$

特别地, $\mathbb{E}[X^2] = \int_0^\infty 2t\mathbb{P}(X > t)\mathrm{d}t$。

由引理 2 可得

$$\mathbb{E}[N(t)] = \sum_{n=1}^\infty \mathbb{P}\{N(t) \geqslant n\} = \sum_{n=1}^\infty \mathbb{P}\{T_n \leqslant t\} \tag{3}$$

对 Poisson 过程, $\mathbb{E}[N(t)] = \lambda t$。而在一般的更新过程中, 得不到关于 $\mathbb{E}[N(t)]$ 的明确的解析表达式。此时可考虑上下界。

3. $\mathbb{E}[N(t)]$ 的上下界

首先介绍 Wald 方程。设 $x_1, x_2, \cdots$ 是独立同分布的随机变量, $\mu = \mathbb{E}[x_i]$。记 $S_n = x_1 + \cdots + x_n$, 设 N 是停时, 即"在时刻 n 终止"可由 $S_1, S_2, \cdots, S_n$ 的取值确定。

引理 4 (Wald 方程) 设 $\mathbb{E}(N) < \infty$, 那么 $\mathbb{E}[S_N] = \mu\mathbb{E}[N]$。

证明 引入示性函数

$$1_N(m) = \begin{cases} 1, & m \leqslant N \\ 0, & 其他 \end{cases}$$

那么

$$S_N = \sum_{m=1}^{\infty} x_m 1_N(m)$$

注意到 $\{N \geqslant m\}^c = \bigcup_{k=1}^{m-1} \{N = k\}$, 因为 N 是停时, 所以 $\{N \geqslant m\}^c$ 和 $\{N \geqslant m\}$ 可由 $x_1, x_2, \cdots, x_{m-1}$ 确定, 即 $1_N(m)$ 和 x_m 是独立的, 从而

$$\mathbb{E}[S_N] = \sum_{m=1}^{\infty} \mathbb{E}x_m \mathbb{E}[1_N(m)] = \mu \sum_{m=1}^{\infty} \mathbb{P}\{N \geqslant m\} = \mu\mathbb{E}[N]$$

设更新过程的 $\{t_n, n = 1, 2, \cdots\}$ 独立同分布, $\mathbb{E}(t_i) = \mu$。

(i) 现在应用引理 4 求 $\mathbb{E}[N(t)]$ 的下界。首先注意到

$$N = \min\{n : T_n > t\} = N(t) + 1$$

是停时。这是因为

$$\{N = n\} = \{N(t) = n - 1\} = \{T_{n-1} \leqslant t < T_n\}$$

可由 $T_1, T_2, \cdots, T_n$ 的取值确定。

由引理 4 可得

$$\mathbb{E}[T_{N(t)+1}] = \mu\mathbb{E}[N(t) + 1]$$

因为 $T_{N(t)+1} > t$, 最后可得到 $\mathbb{E}[N(t)]$ 的下界如下

$$\mathbb{E}[N(t)] + 1 > \frac{t}{\mu} \tag{4}$$

(ii) 为求 $\mathbb{E}[N(t)]$ 的上界, 设 $t_i \leqslant M$, 那么 $T_{N(t)+1} \leqslant t + M$, 从而

$$\mathbb{E}[N(t) + 1] = \frac{1}{\mu}\mathbb{E}[T_{N(t)+1}] \leqslant \frac{t + M}{\mu} \tag{5}$$

4. 一些特殊情形下 $\mathbb{E}[N(t)]$ 的解析表达式

设更新过程的 $\{t_n, n=1,2,\cdots\}$ 独立同分布, t_i 的分布函数为 $F(t)$, 密度函数为 $f(t)$, $\mathbb{E}(t_i)=\mu$。

以下讨论的出发点是 (3) 式。记 T_n 的分布函数为 $K_n(t)$, 它是分布函数 $F(t)$ 的 n 重卷积。另外, 记 $H(t)=\mathbb{E}[N(t)]$, 那么 (3) 式简写为

$$H(t)=\sum_{n=1}^{\infty}K_n(t) \tag{6}$$

为求解 (6) 式, 引入 Laplace 变换。函数 $f(t):[0,\infty)\to[0,\infty)$ 的 Laplace 变换定义为

$$f^*(s)=\int_0^{\infty}\mathrm{e}^{-st}f(t)\mathrm{d}t, s>0 \tag{7}$$

在以下讨论中用到 Laplace 变换的以下两条性质:

(i) $F(t)=\int_0^t f(s)\mathrm{d}s$ 的 Laplace 变换为

$$F^*(s)=\int_0^{\infty}\mathrm{e}^{-st}F(t)\mathrm{d}t=\frac{1}{s}f^*(s)$$

(ii) 设 $h(t)=\int_0^t f(s)g(t-s)\mathrm{d}s$, 那么

$$h^*(s)=f^*(s)g^*(s)$$

对 (6) 式两边取 Laplace 变换, 利用关于更新过程的假设, 以及上面两条性质, 最后得到

$$H^*(s)=\sum_{n=1}^{\infty}\frac{1}{s}[f^*(s)]^n=\frac{f^*(s)}{s[1-f^*(s)]} \tag{8}$$

在一些特殊的情形中, 由 (8) 式求 Laplace 变换的逆变换, 就得 $H(t)=\mathbb{E}[N(t)]$。

例 3 设 $f(t)=4t\mathrm{e}^{-2t}$, 求 $\mathbb{E}[N(t)]$。

此时 $f^*(s)=\dfrac{4}{(s+2)^2}$, 从而

$$H^*(s)=\frac{f^*(s)}{s[1-f^*(s)]}=\frac{4}{s^2(s+4)}$$

求 Laplace 变换的逆变换, 得到

$$\mathbb{E}[N(t)]=t-0.25(1-\mathrm{e}^{-4t})$$

当 $t\to\infty$ 时, $\mathbb{E}[N(t)]\to t-\dfrac{1}{4}$。

例 4　设 $f(t)=\frac{1}{2}e^{-t}+\frac{1}{2}(3e^{-3t})$, 求 $\mathbb{E}[N(t)]$。

此时 $f^*(s)=\frac{1}{2}\frac{1}{s+1}+\frac{1}{2}\frac{3}{s+3}$, 从而

$$H^*(s)=\frac{f^*(s)}{s[1-f^*(s)]}=\frac{2s+3}{s^2(s+2)}$$

求 Laplace 变换的逆变换, 得到

$$\mathbb{E}[N(t)]=\frac{3t}{2}-\frac{1}{4}(e^{-2t}-1)$$

当 $t\to\infty$时, $\mathbb{E}[N(t)]\to\frac{3t}{2}+\frac{1}{4}$。

§2　关于更新次数的极限

对一般的更新过程, 求 $\mathbb{E}[N(t)]$ 的解析表达式是困难的, 此时可转而考虑更新次数 $N(t)$ 的极限。

1. 强大数律

定理 1　记 $\mu=\mathbb{E}t_i$ 为平均更新时间, 那么当 $t\to\infty$ 时, 下式几乎处处成立

$$\frac{N(t)}{t}\to\frac{1}{\mu}\tag{9}$$

或等价地, 当 $t\to\infty$ 时, $N(t)\approx\frac{t}{\mu}$。

证明　为证明定理 1, 引用概率论中的强大数律, 即对独立同分布的随机变量 $x_1,x_2,\cdots$, $\mathbb{E}x_i=\mu$。记 $S_n=x_1+\cdots+x_n$, 那么当 $n\to\infty$ 时, $S_n/n\to\mu$ 几乎处处成立。

令 $x_i=t_i$, 从而 $S_n=T_n$, 应用强大数律, 可得当 $n\to\infty$ 时, $T_n/n\to\mu$ 几乎处处成立。

由定义,

$$T_{N(t)}\leqslant t<T_{N(t)+1}$$

从而

$$\frac{T_{N(t)}}{N(t)}\leqslant\frac{t}{N(t)}<\frac{T_{N(t)+1}}{N(t)}=\frac{T_{N(t)+1}}{N(t)+1}\frac{N(t)+1}{N(t)}\tag{10}$$

应用强大数律, 当 $t\to\infty$ 时, (10) 式两边极限都是 μ。由此得到定理结论。

下面给出另一定理。

定理 2 记 $\mu = \mathbb{E}t_i$ 为平均更新时间, 那么当 $t \to \infty$ 时

$$\frac{\mathbb{E}[N(t)]}{t} \to \frac{1}{\mu} \tag{11}$$

证明 应用 § 1 给出的 $\mathbb{E}[N(t)]$ 的上下界。一方面

$$\mathbb{E}[N(t)] + 1 \geqslant \frac{t}{\mu} \tag{12}$$

由 (12) 式可得

$$\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} \geqslant \frac{1}{\mu} \tag{13}$$

另一方面, 对正数 M, 设 $\bar{t}_i = \min\{t_i, M\}$, $\mu_M = \mathbb{E}[\bar{t}_i]$, $\overline{T}_n = \bar{t}_1 + \cdots + \bar{t}_n$。因为 $\bar{t}_i \leqslant t_i$, 所以 $\overline{T}_n \leqslant T_n$, $\overline{N}(t) \geqslant N(t)$。

$$\mathbb{E}[N(t)] < \mathbb{E}[\overline{N}(t) + 1] = \frac{\mathbb{E}[\overline{T}_{\overline{N}(t)+1}]}{\mu_M} \leqslant \frac{t + M}{\mu_M} \tag{14}$$

由 (14) 式可得

$$\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} \leqslant \frac{1}{\mu_M} \tag{15}$$

最后令 $M \to \infty$, 注意到 $\mu_M = \mathbb{E}[\bar{t}_i] \to \mathbb{E}[t_i] = \mu$, 由 (15) 式即得

$$\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} \leqslant \frac{1}{\mu} \tag{16}$$

综合 (13) 式和 (16) 式, 即证定理。

注意到以上两个定理所涉及的两种极限是不同的, 定理 2 不能由定理 1 推出。下面给出一个例子来说明这一点。

例 5 设 U 是 (0,1) 上的均匀分布变量, 定义

$$Y_t = \begin{cases} c + t, & U \leqslant 1/t \\ c, & U > 1/t \end{cases}$$

那么当 $t \to \infty$ 时, $Y_t \to c$ 几乎处处成立。但是当 $t \geqslant 1$ 时,

$$\mathbb{E}[Y_t] = c\mathbb{P}\{U > 1/t\} + (c + t)\mathbb{P}\{U \leqslant 1/t\} = c + 1$$

所以 $\mathbb{E}Y_t = c + 1 > c$。

2. 更新过程的概念推广

设在第 i 次更新时刻，奖励额为 r_i。进一步，设 $\{(r_i,t_i), i=1,2,\cdots\}$ 是独立同分布的。记 $R(t)=\sum_{i=1}^{N(t)} r_i$ 为到时刻 t 时总的奖励额。

定理 3 下列极限几乎处处存在

$$\frac{R(t)}{t} \to \frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]},\ t\to\infty$$

证明 由定理 1 即得

$$\frac{R(t)}{t}=\left(\frac{1}{N(t)}\sum_{i=1}^{N(t)} r_i\right)\frac{N(t)}{t}\to \mathbb{E}[r_i]\frac{1}{\mathbb{E}[t_i]}$$

例 6 设汽车模型的寿命随机变量的密度函数为 h。当汽车模型失效或经过 T 段时间，需要更新模型。新的模型的成本为 A，而当模型在 T 段时间之前失效时，额外维修成本是 B。从长远来看，每单位时间内的平均成本是多少？

在此例中，第 i 个周期的长度记为 t_i，那么

$$\mathbb{E}[t_i]=\int_0^T th(t)\mathrm{d}t+T\int_T^\infty h(t)\mathrm{d}t$$

第 i 个周期的成本为 r_i，

$$\mathbb{E}[r_i]=A+B\int_0^T h(t)\mathrm{d}t$$

由定理 3 即得每单位时间内的平均成本如下

$$\frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]}=\frac{A+B\displaystyle\int_0^T h(t)\mathrm{d}t}{\displaystyle\int_0^T th(t)\mathrm{d}t+T\int_T^\infty h(t)\mathrm{d}t}$$

例 7 考虑 §1 中的机器维修的例子。记 $\mathbb{E}s_i=\mu_1$，$\mathbb{E}u_i=\mu_2$。从长远看，更新过程处于工作状态的时间比例为 $\dfrac{\mu_1}{\mu_1+\mu_2}$。

为说明这一结论，记 $t_i=s_i+u_i$ 为第 i 个周期的持续时间，$r_i=s_i$。如果忽略在时刻 t 时仍然持续的那个周期，那么 $R(t)=\sum_{i=1}^{N(t)} r_i$ 就是到时刻 t 时更新过程处于工作状态的时间。由定理 3 即得所求的时间比例为

$$\frac{R(t)}{t}\to\frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]}=\frac{\mu_1}{\mu_1+\mu_2}$$

§3 年龄与剩余寿命

本节要说明当时间趋于无穷时, 更新过程将达到平衡状态, 这与 Markov 链的性质相类似。在本节的论述中, § 2 节介绍的推广的更新过程概念起着关键的作用。

设 $\{t_i, i = 1, 2, \cdots\}$ 是独立同分布变量, $T_n = t_1 + \cdots + t_n$, $N(t) = \max\{n : T_n \leqslant t\}$。引入符号

$$A(t) = t - T_{N(t)}, Z(t) = T_{N(t)+1} - t$$

那么 $A(t)$ 表示包含时刻 t 的更新周期已经经历的时间, $Z(t)$ 表示包含时刻 t 的更新周期的未来剩余时间。在灯管寿命模型中, $A(t)$ 就是在时刻 t 仍在使用的灯管的年龄, 而 $Z(t)$ 就是灯管的剩余寿命。

1. 平均剩余寿命

为考虑如下平均剩余寿命

$$\frac{1}{t}\int_0^t Z(s)\mathrm{d}s$$

首先注意到当 $T_{i-1} < s < T_i$ 时, $Z(s) = T_i - s$, 从而

$$\int_{T_{i-1}}^{T_i} Z(s)\mathrm{d}s = \int_0^{t_i} r\mathrm{d}r = \frac{t_i^2}{2} \tag{17}$$

在 (17) 式的推导中用到变量替换 $r = T_i - s$。

忽略最后的不完全周期 $[T_{N(t)}, t]$, 可得

$$\int_0^t Z(s)\mathrm{d}s \approx \sum_{i=1}^{N(t)} \frac{t_i^2}{2}$$

由 § 2 节定理 3 可得下列极限几乎处处存在

$$\frac{1}{t}\int_0^t Z(s)\mathrm{d}s \to \frac{\mathbb{E}[t_i^2/2]}{\mathbb{E}[t_i]} \tag{18}$$

2. 平均年龄

应用与前面类似的方法, 做变量替换 $s = T_{i-1} + r$, 可得

$$\int_{T_{i-1}}^{T_i} A(s)\mathrm{d}s = \int_0^{t_i} r\mathrm{d}r = \frac{t_i^2}{2} \tag{19}$$

忽略最后的不完全周期 $[T_{N(t)}, t]$, 可得

$$\int_0^t A(s)\mathrm{d}s \approx \sum_{i=1}^{N(t)} \frac{t_i^2}{2}$$

由 § 2 节定理 3 可得下列极限几乎处处存在

$$\frac{1}{t}\int_0^t A(s)\mathrm{d}s \to \frac{\mathbb{E}[t_i^2/2]}{\mathbb{E}[t_i]} \tag{20}$$

3. 时刻 t 之前的平均寿命

记 $L(t)=A(t)+Z(t)$ 为包含时刻 t 的更新周期时间。到时刻 t 之前平均寿命为

$$\frac{1}{t}\int_0^t L(s)\mathrm{d}s$$

综合前面的内容, 可知

$$\frac{1}{t}\int_0^t L(s)\mathrm{d}s \to \frac{\mathbb{E}[t_i^2]}{\mathbb{E}[t_i]} \tag{21}$$

(21) 式是一个很惊奇的结论, 这是因为一般地, $\mathrm{Var}(t_i)=\mathbb{E}[t_i^2]-(\mathbb{E}[t_i])^2>0$, 从而 (21) 式右边大于 $\mathbb{E}t_i$。但是, 根据大数律, 前 n 个更新周期的平均时间是 $\mathbb{E}t_i$。

(21) 式表明: 对时刻 t 之前的更新周期取平均值会出现向上的偏差。这是因为对 t 之前的更新周期, 如果有些更新周期的时间相同, 设出现 m 次, 那么将统计 m 次。为简要说明, 设寿命变量取 k 个值 $l_1<l_2<\cdots<l_k$, 相应的概率分别为 $p_1,p_2,\cdots,p_k$。

考虑一个推广的更新过程, 只有当 $t_i=l_j$ 时, $r_i=1$, 那么到时刻 t 前出现的寿命为 l_j 的更新周期有 $p_jN(t)$ 个, 总的寿命为 $l_jp_jN(t)$。从而出现寿命为 l_j 的时间比例为

$$\frac{l_jp_jN(t)}{t} \to \frac{l_jp_j}{\mathbb{E}[t_i]}.$$

最后可求得期望值

$$\sum_j l_j\frac{l_jp_j}{\mathbb{E}[t_i]}=\frac{\mathbb{E}[t_i^2]}{\mathbb{E}[t_i]}$$

4. 剩余寿命的极限分布

已知 $Z(t)=T_{N(t)+1}-t$, 对 $c>0$, 引入示性函数

$$1_c(t)=\begin{cases}1, & Z(t)\leqslant c\\ 0, & 其他\end{cases}$$

当 $t\to\infty$ 时, 为求 $\mathbb{P}(Z(t)\leqslant c)$ 的极限概率, 转而考虑

$$\frac{1}{t}\int_0^t 1_c(s)\mathrm{d}s$$

为此, 先考虑 $\int_{T_{i-1}}^{T_i} 1_c(s)\mathrm{d}s$。下面验证

$$\int_{T_{i-1}}^{T_i} 1_c(s)\mathrm{d}s = \min\{t_i, c\} \tag{22}$$

(i) 当 $t_i \geqslant c$ 时, $1_c(s) = 1$ 只在 $T_i - c \leqslant s \leqslant T_i$ 上成立, 从而 $\int_{T_{i-1}}^{T_i} 1_c(s)\mathrm{d}s = c$。

(ii) 当 $t_i \leqslant c$ 时, $1_c(s) = 1$ 在整个区间上 $T_{i-1} \leqslant s \leqslant T_i$ 成立, 从而 $\int_{T_{i-1}}^{T_i} 1_c(s)\mathrm{d}s = t_i$。忽略最后的不完全周期 $[T_{N(t)}, t]$, 可得

$$\int_0^t 1_c(s)\mathrm{d}s \approx \sum_{i=1}^{N(t)} \min\{t_i, c\} \tag{23}$$

(23) 式右边是一个推广的更新过程, $r_i = \min\{t_i, c\}$。由 § 2 节定理 3 可得

$$\frac{1}{t}\int_0^t 1_c(s)\mathrm{d}s \to \frac{\mathbb{E}[\min\{t_i, c\}]}{\mathbb{E}[t_i]} \tag{24}$$

由 (24) 式可得, 当 $t \to \infty$ 时,

$$\mathbb{P}\{Z(t) \leqslant c\} \to \frac{\mathbb{E}[\min\{t_i, c\}]}{\mathbb{E}[t_i]} \tag{25}$$

5. 年龄的极限分布

已知 $A(t) = t - T_{N(t)}$, 对 $c > 0$, 引入示性函数

$$1_c(t) = \begin{cases} 1, & A(t) \leqslant c \\ 0, & \text{其他} \end{cases}$$

应用与前面完全类似的方法, 可得到相同的结论, 即当 $t \to \infty$ 时,

$$\mathbb{P}\{A(t) \leqslant c\} \to \frac{\mathbb{E}[\min\{t_i, c\}]}{\mathbb{E}[t_i]} \tag{26}$$

6. 均衡更新过程

注意到

$$\mathbb{E}[\min\{t_i, c\}] = \int_0^\infty \mathbb{P}\{\min\{t_i, c\} > t\}\mathrm{d}t = \int_0^c \mathbb{P}\{t_i > t\}\mathrm{d}t$$

所以, 如记

$$G(c) = \frac{\mathbb{E}[\min\{t_i, c\}]}{\mathbb{E}[t_i]}$$

那么

$$G(c)=\frac{\int_0^c \mathbb{P}\{t_i>t\}\mathrm{d}t}{\mathbb{E}[t_i]} \tag{27}$$

容易验证 $G(c)$ 是一个分布函数, 相应的密度函数为

$$g(c)=G'(c)=\frac{\mathbb{P}\{t_i>c\}}{\mathbb{E}[t_i]} \tag{28}$$

例 8　设 t_i 是参数为 λ 的指数变量, 那么

$$\frac{\mathbb{P}\{t_i>c\}}{\mathbb{E}[t_i]}=\frac{\mathrm{e}^{-\lambda c}}{1/\lambda}=\lambda\mathrm{e}^{-\lambda c}$$

如果考虑到指数分布的无记忆性, 那么剩余寿命 $Z(t)$ 的分布总是参数为 λ 的 Poisson 分布, 这也验证了本例的结论。

例 9　设 t_i 是 $(0,b)$ 区间上的均匀分布变量, 那么

$$\frac{\mathbb{P}\{t_i>c\}}{\mathbb{E}[t_i]}=\frac{(b-c)/b}{b/2}=\frac{2(b-c)}{b^2}$$

即极限密度函数是线性函数, 密度函数当 $c=0$ 时取值为 $2/b$, 而当 $c=b$ 取值为 0。

一般地, 设 t_i 的分布函数为 F。下面的定理说明 Poisson 过程是唯一的更新过程, 使得 $G=F$。

定理 4　假设 $G=F$, 那么存在 $\lambda>0$, 使得 $F(x)=1-\mathrm{e}^{-\lambda x}$。

证明　记 $H(c)=1-G(c)$, 两边取导数, 那么由 (28) 式可得 $H'(c)=-\lambda\mathbb{P}[t_i>c]=-\lambda(1-F(c))$, 这里记 $\lambda=1/\mathbb{E}[t_i]$。再由假设$G=F$, 可得 $H'(c)=-\lambda(1-G(c))=-\lambda H(c)$。微分方程 $H'(c)=-\lambda H(c)$ 的初值条件为 $H(0)=1$。求解此微分方程, 即得 $H(c)=\mathrm{e}^{-\lambda c}$。

最后可以总结一下更新过程的极限行为。设更新过程已经持续了很长的 t 段时间, 记 $\{T_1',T_2',\cdots\}$ 为从时刻 t 开始, 以后的更新出现的时间。再记 $t_1'=T_1'$, $t_k'=T_k'-T_{k-1}'$, $k\geqslant 2$。那么当 t 很大时, t_1' 的分布近似地为 G。由此引出如下定义。

定义　设随机变量 $\{t_1,t_2,t_3,\cdots\}$ 是相互独立的, 其中 $\{t_2,t_3,\cdots\}$ 独立同分布, 分布函数为 F, 即 $\mathbb{P}\{t_i>c\}=1-F(c)$, t_1 的分布函数为由 (28) 式定义的 G, 那么称 $\{t_1,t_2,t_3,\cdots\}$ 构成的更新过程为均衡更新过程。

对均衡更新过程, 类似地有 $N(t)$ 的定义。记 $H_e(t) = \mathbb{E}[N(t)]$, 应用 § 1 节中 Laplace 变换方法, 对那里的推导加以调整, 可以得到

$$H_e^*(s) = \frac{1}{\mu}\frac{1}{s^2} \tag{29}$$

其中 $\mu = \mathbb{E}t_2$, 从而

$$H_e(t) = \frac{t}{\mu} \tag{30}$$

(30) 式表明, 在均衡更新过程中, $\mathbb{E}[N(t)]$ 有很简单的形式。

进一步, 对均衡更新过程, 下面要证明: 对给定的正数 t, $T_{N(t)+1} - t$ 的分布不依赖于 t, 而且它的分布正是 G。

按定义, $T_{N(t)} \leqslant t < T_{N(t)+1}$, 根据 $N(t)$ 的取值, 由全概率公式, 可得当 $x > 0$ 时,

$$\begin{aligned}
\mathbb{P}\{T_{N(t)+1} - t \leqslant x\} &= \sum_{n=0}^{\infty} \mathbb{P}\{T_n \leqslant t < T_{n+1} \leqslant t + x\} \\
&= \mathbb{P}\{t < T_1 \leqslant t + x\} + \sum_{n=1}^{\infty} \mathbb{P}\{T_n \leqslant t < T_{n+1} \leqslant t + x\} \\
&= \mathbb{P}\{t < T_1 \leqslant t + x\} + \sum_{n=1}^{\infty} \int_0^t [F(t + x - y) - \\
&\quad F(t - y)]\mathrm{d}(G * F^{(n-1)*})(y) \\
&= G(t + x) - G(t) + \frac{1}{\mu}\int_0^t [F(t + x - y) - F(t - y)]\mathrm{d}y \\
&= G(x)
\end{aligned}$$

在上面的推导中, 第四个等式应用了 (30) 式, 而最后一个等式应用 G 的定义。

由上面的结论, 容易理解, 对均衡更新过程, 平稳增量性是成立的, 即对 $t_0 < t_1 < t_2 < \cdots < t_n$, 以下随机向量的联合分布不依赖于 t

$$\{N(t_1 + t) - N(t_0 + t), N(t_2 + t) - N(t_1 + t), \cdots, N(t_n + t) - N(t_{n-1} + t)\}$$

§4 更新方程简介

直观地讲, 更新方程出现于更新过程中。在讨论涉及保险风险过程的破产概率问题时, 更新方程会经常出现。这里仅给出一个简单的介绍。

1. 定义

从一个简单的问题出发。由 § 1 节的 (3) 式,

$$H(t) = \mathbb{E}[N(t)] = \sum_{n=1}^{\infty} \mathbb{P}\{T_n \leqslant t\} = \sum_{n=1}^{\infty} F^{*n}(t) \tag{31}$$

其中 F 表示随机变量 t_i 的分布函数。

由 (31) 式, 即得

$$\begin{aligned} H(t) &= \sum_{n=1}^{\infty} F^{*n}(t) = F(t) + F * \left(\sum_{n=1}^{\infty} F^{*n}\right)(t) \\ &= F(t) + F * H(t) \\ &= F(t) + \int_0^t H(t-s)\mathrm{d}F(s) \end{aligned} \tag{32}$$

(32) 式就是一个简单的更新方程。

一般地, 设 F 是定义于 $[0,\infty)$ 上的非负单增函数, 而且 $F(\infty) < \infty$, 而 z 是定义于 $[0,\infty)$ 上的非负函数, 那么称以下形式的方程为更新方程

$$g(x) = z(x) + \int_0^x g(x-y)\mathrm{d}F(y) \tag{33}$$

比较 (32) 式和 (33) 式, 可知当 $z = F$ 时, 方程 (33) 的解就是 $g = H$。因此 (32) 式是一个特殊的更新过程。

在 $z(x)$ 是局部有界的条件下, 方程 (33) 存在唯一的正解。为给出解, 引入

$$H_0(t) = \sum_{n=0}^{\infty} F^{*n}(t)$$

当 $t \geqslant 0$ 时, $H_0(t) = 1 + H(t)$。这是因为 F^{*0} 是对应于空和(即零)的分布函数, 所以

$$F^{*0}(t) = \begin{cases} 1, & t \geqslant 0 \\ 0, & t < 0 \end{cases}$$

F^{*0} 的导数是 δ 函数, 它的含义就是对任意定义于 $\mathbb{R}$ 上的函数 f,

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}F^{*0}(t) = f(0)$$

与 $H(t)$ 相比, $H_0(t)$ 把时刻 0 也视为一个更新点。

直观上, 在方程 (33) 式右边叠代代入, 可得到如下形式的唯一解。关于此结论的严格证明这里略去。

$$g(x) = \sum_{n=0}^{\infty} \int_0^x z(x-y)\mathrm{d}F^{*n}(y) = \int_0^x z(x-y)\mathrm{d}H_0(y) \tag{34}$$

需要指出, 在 (34) 式右边的积分要在广义积分下理解, 即应用 δ 函数的定义。

例 10 在 (33) 式中取 $z=1$, 那么更新方程 $g(x)=1+\int_0^x g(x-y)\mathrm{d}F(y)$ 的唯一解为 $g(x)=H_0(x)$。

例 11 设 F 为分布函数, 对应的期望 $\mu<\infty$。在 (33) 式中取 $z=G$, 那么更新方程的解为 $g(x)=x/\mu$。这是因为, 由 (34) 式, 以及均衡更新过程的性质, 即得

$$g(x)=\sum_{n=0}^{\infty}\int_0^x G(x-y)\mathrm{d}F^{*n}(y)=H_e(x)=\frac{x}{\mu}$$

2. 解的渐近表示

为便于讨论, 进一步假设 $F'=f$, 那么 (33) 式就变为

$$g(x)=z(x)+\int_0^x g(x-y)f(y)\mathrm{d}y \tag{35}$$

为考虑方程 (35) 的解当 $x\to\infty$ 时的渐近性, 选择实数 R, 使得

$$\hat{f}(R)=\int_0^{\infty}\mathrm{e}^{Rx}f(x)\mathrm{d}x=1$$

上面选择的实数 R 如果存在, 那么它是唯一的, 这是因为函数 $\hat{f}$ 是单增的,

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\hat{f}(\xi)=\int_0^{\infty}\mathrm{e}^{\xi x}xf(x)\mathrm{d}x>0$$

显然, 当 f 是概率密度函数时, $R=0$。

更新方程的解的渐近性结论表明, 对比较正则的函数 z, 当 $x\to\infty$ 时,

$$\mathrm{e}^{Rx}g(x)\to\frac{\displaystyle\int_0^{\infty}\mathrm{e}^{Ry}z(y)\mathrm{d}y}{\displaystyle\int_0^{\infty}y\mathrm{e}^{Ry}f(y)\mathrm{d}y} \tag{36}$$

有时候, 解的渐近性也表示为如下形式

$$g(x)\to C\mathrm{e}^{-Rx}$$

其中 C 表示 (36) 式右边确定的常数。

参考文献

Asmussen, S. (2000). Ruin Probabilities. World Scientific, Singapore.

Baxter, M., Rennie, A. (1996). Financial Calculus. Cambridge University Press. Reprinted 1998.

Bertoin, J. (1999). Some elements on Lévy processes. Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VI. Prepublication No. 485.

Bowers, N., Gerber, H., Hickman, J., Jones, D., Nesbitt, C. (1997). Actuarial Mathematics (Second Edition). Society of Actuaries, Illinois, USA.

Cairns, A., Dickson, D., Macdonald, A., Waters, H., Willder, M. (2000). Stochastic processes: learning the language. The University of Melbourne, Centre for Actuarial Studies, Research Paper Series No. 76.

Cox, J., Ingersoll, J., Ross, S. (1985). A theory of the term structure of interest rates. Econometrica, 53, 385-405.

Dickson, D. (2001). Modern landmarks in actuarial science. Australian Actuarial Journal, 7 (1), 217-245.

Drekic, S., Dickson, D., Stanford, D., Willmot, G. (2004). On the distribution of the deficit at ruin when claims are phase-type. Scandinavian Actuarial Journal, 105-120.

Durrett, R. (1999). Essentials of Stochastic Processes. Springer, New York.

Etheridge, A. (2002). A Course in Financial Calculus. Cambridge University Press.

Jacod, J., Protter, P. (1999). Probability Essentials. Springer.

Karatzas, I., Shreve, S. (1988). Brownian Motion and Stochastic Calculus. Springer-Verlag. Second Edition 1991.

Klebaner, F. (1998). Introduction to Stochastic Calculus with Applications. Imperial College Press. Reprinted 2001.

Mikosch, T. (2003). Non-life Insurance Mathematics: an Introduction with Stochastic Processes. Springer.

Mörters, P. (2005). Martingale Theory. Lecture Notes. Available from http://www.bath.ac.uk

Norberg, R. (1995a). Differential equations for moments of present values in life insurance. Insurance: Mathematics and Economics 17, 171-180.

Norberg, R. (1995b). Stochastic calculus in actuarial science. Surveys in Industrial and Applied Mathematics, 2 (5), 1-23.

Norberg, R. (1995c). A time-continuous Markov chain interest model with applications to insurance. Applied Stochastic Models and Data Analysis, 11, 245-256.

Norberg, R. (2003). The Markov chain market. Astin Bulletin, 33 (2), 265-287.

Norberg, R. (2004). Vasicek beyond the normal. Mathematical Finance, 14 (4), 585-604.

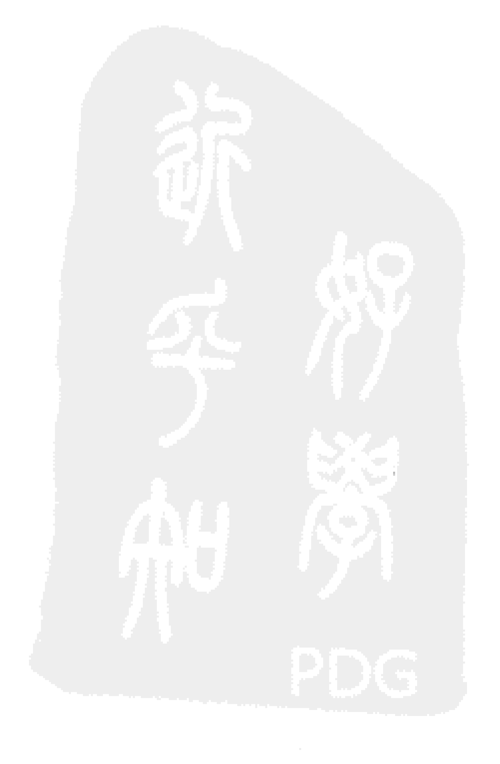
Norberg, R. (2004). Risk and Stochastics in Life Insurance. Lecture Notes.

Rolski, T., Schmidli, H., Schmidt, V., Teugels, J. (1999). Stochastic Processes for Insurance and Finance. New York: John Wiley & Sons.

Ross, S. (2003). Introduction to Probability Models (Eighth Edition). Academic Press.

Teugels, J., Sundt, B. (editors-in-chief) (2004). Encyclopedia of Actuarial Science. John Wiley.

Vasicek, O. (1977). An equilibrium characterization of the term structure. Journal of Financial Economics, 5, 177-188.



名词索引





